



UNIVERSIDAD NACIONAL DE LA PLATA

FACULTAD DE CIENCIAS EXACTAS

DEPARTAMENTO DE CIENCIAS BIOLÓGICAS

Trabajo de Tesis Doctoral:

**BÚSQUEDA DE NUEVOS FÁRMACOS ANTIHELMÍNTICOS
BASADA EN EL ANÁLISIS ESTRUCTURA-FUNCIÓN DE
PROTEÍNAS QUE UNEN ÁCIDOS GRASOS (FABPs) DE
CESTODOS MEDIANTE TÉCNICAS BIOINFORMÁTICAS Y
QUIMIOINFORMÁTICAS**

Tesista: Lic. Santiago Rodríguez

Directora: Dra. Gisela Raquel Franchini

Codirector: Dr. Alan Talevi

Año: 2024

A mis héroes de ayer, hoy y siempre

AGRADECIMIENTOS

En primer lugar, quisiera agradecer a mis directores: **Dra. Gisela R. Franchini** y **Dr. Alan Talevi**, como así también a la **Dra. Betina Córscico**, por haberme abierto las puertas del **INIBIOLP**, del **LIDeB**, y del mundo.

A **los jurados**, por haberse tomado el trabajo de leer y evaluar esta tesis doctoral.

A **CONICET**, la **UNLP**, la **Facultad de Ciencias Exactas** y al **Colegio Monseñor Alberti** por mi formación tanto académica como general.

Al consorcio **REFRACT**, por el financiamiento económico otorgado para las pasantías realizadas en Francia y Suiza, en una aventura que me cambió la vida.

Al **rugby** en general, y al **Club Los Tilos** en particular, por inculcarme los valores de la amistad, compañerismo, compromiso y lealtad, los cuales llevo conmigo todos los días.

Al **Dr. Pau Bernadó**, por haberme recibido en su grupo en el *Centre de Biologie Structurale (CBS)* y al **Dr. Carlos Elena-Real** por su apoyo tanto dentro como fuera del **CBS**, en un momento en el cual necesitaba a un amigo. De la misma manera, quisiera agradecer al resto de mis colegas del **CBS**, por ser mi familia mientras yo estaba en Francia.

A la **Dra. María Anisimova** y su grupo, por haberme recibido en el *Institute of Computational Life Sciences (ICLS-ZHAW)*, en la ciudad de Wädenswil, Suiza. Allí tuve la dicha de conocer al **Lic. Max Verbiest** y a su novia, la **Lic. Alice Driessen**, quienes de forma desinteresada estuvieron conmigo en un momento personal muy difícil y hoy tenemos una gran amistad.

A **mis compañeros** del **INIBIOLP** y del **LIDeB**, por los momentos compartidos, en los cuales intenté aprender un poco de cada uno de ustedes.

A **mis amigos**, por estar en las buenas y en las malas mucho más.

A **mis tíos y primos**, por sus buenas vibras siempre.

A mi pareja, **Gisele**, gracias por creer.

A mi hermano **Andrés**, por escucharme y hacer las veces de hermano mayor. Gracias también por tu generosa donación de placa gráfica y demás componentes al **LIDeB**. Gracias **Agustina** por hacer de él una persona aún mejor.

A mis padres, **Mariano y Edith**, gracias por ser mi espejo. No hubiera llegado hasta acá si no hubiera sido por ustedes.

A mis abuelos, **Oscar, Osvaldo, Nine, Hebe, Julia y Selva**, gracias por la niñez, probablemente la envidia de cualquier niño hoy en día.

Muchas gracias a todos.

ÍNDICE

ACRÓNIMOS	11
ESTADÍAS, PUBLICACIONES Y CONGRESOS	16
OBJETIVO GENERAL	19
OBJETIVOS ESPECÍFICOS.....	19
CAPÍTULO 1: INTRODUCCIÓN	21
1.1 ENFERMEDADES TROPICALES DESATENDIDAS.....	21
1.2 EQUINOCOCOSIS, TENIASIS Y CISTICERCOSIS.....	22
1.2.1 GENERALIDADES.....	22
1.2.2 <i>E. GRANULOSUS</i> , <i>E. MULTILOCULARIS</i> Y <i>T. SOLIUM</i>	24
1.2.3 CICLO DE VIDA DE LOS PARÁSITOS.....	24
1.2.4 EPIDEMIOLOGÍA Y DISTRIBUCIÓN.....	27
1.2.5 MANIFESTACIONES CLÍNICAS.....	29
1.2.6 MEDICAMENTOS DISPONIBLES.....	30
1.2.7 AVANCES HACIA NUEVOS MEDICAMENTOS.....	32
1.3 BLANCOS MOLECULARES EN CESTODOS.....	34
1.3.1 INHIBIDORES DE LA RESPIRACIÓN ANAERÓBICA.....	35
1.3.2 TIORREDOXINA GLUTATIÓN REDUCTASA.....	35
1.3.3 PROTEÍNA ADN DAÑO-INDUCIBLE 1 DE <i>E. MULTILOCULARIS</i>	36
1.4 PROTEÍNAS DEL METABOLISMO LIPÍDICO.....	37
1.4.1 GENERALIDADES.....	37
1.4.2 METABOLISMO DE LÍPIDOS.....	38
1.4.3 FABPs.....	40
1.5 ELECCIÓN DEL BLANCO MOLECULAR.....	41
REFERENCIAS	42
CAPÍTULO 2: DESCUBRIMIENTO DE FÁRMACOS	52
2.1 INTRODUCCIÓN.....	52
2.1.1 DESCUBRIMIENTO DE FÁRMACOS PARA ETDs.....	53
2.2 CRIBADO VIRTUAL.....	55
2.2.1 MÉTODOS BASADOS EN EL BLANCO MOLECULAR.....	57
2.2.1.1 MODELADO POR APRENDIZAJE AUTOMÁTICO.....	60
2.2.1.2 DOCKING MOLECULAR.....	62
2.2.1.3 DINÁMICA MOLECULAR.....	65
2.2.2 MÉTODOS BASADOS EN EL LIGANDO.....	67

2.2.2.1 FARMACÓFOROS.....	67
2.2.2.2 SIMILITUD MOLECULAR.....	68
2.2.2.3 DESCRIPTORES MOLECULARES.....	70
2.2.3 MÉTODOS COMBINADOS.....	72
2.3 REPOSICIONAMIENTO DE FÁRMACOS.....	73
2.3.1 GENERALIDADES.....	73
2.3.2 MÉTODOS SISTEMÁTICOS DE REPOSICIONAMIENTO DE FÁRMACOS.....	75
2.3.2.1 ENFOQUES NO COMPUTACIONALES.....	76
2.3.2.2 ENFOQUES COMPUTACIONALES.....	76
2.3.2.2.1 METODOLOGÍAS QUIMIOINFORMÁTICAS.....	76
2.3.2.2.2 METODOLOGÍAS BIOINFORMÁTICAS.....	77
2.3.2.2.3 METODOLOGÍAS BASADAS EN LA LITERATURA.....	78
2.3.3 REPOSICIONAMIENTO DE FÁRMACOS PARA ETDs.....	79
2.3.3.1 CASOS EXITOSOS EN CESTODOS MEDIANTE ENFOQUES COMPUTACIONALES.....	80
2.4 ESTRATEGIA UTILIZADA EN ESTE TRABAJO DE TESIS.....	81
REFERENCIAS.....	83
CAPÍTULO 3: METODOLOGÍA.....	99
3.1 INTRODUCCIÓN.....	99
3.2 MÉTODOS INDIRECTOS. GENERACIÓN DEL MODELO QSAR.....	99
3.2.1 COMPILADO DEL CONJUNTO DE DATOS.....	100
3.2.2 PARTICIÓN DEL CONJUNTO DE DATOS.....	101
3.2.2.1 ESTRATEGIA DE PARTICIÓN UTILIZADA EN ESTE TRABAJO.....	103
3.2.3 CÁLCULO DE DESCRIPTORES MOLECULARES.....	105
3.2.4 MODELADO.....	105
3.2.4.1 SUBESPACIOS ALEATORIOS.....	106
3.2.4.2 SELECCIÓN DE DESCRIPTORES Y GENERACIÓN DE MODELOS.....	106
3.3 VALIDACIÓN DE LOS MODELOS QSAR.....	107
3.3.1 MÉTODOS DE VALIDACIÓN INTERNA.....	108
3.3.2 VALIDACIÓN EXTERNA.....	108
3.3.3 EVALUACIÓN DE LOS MODELOS.....	109

3.3.4	CAMPAÑAS SIMULADAS DE CRIBADO VIRTUAL.....	110
3.3.4.1	QUIMIOTECA SIMULADA LUDE.....	111
3.3.4.2	QUIMIOTECA SIMULADA DUDE-Z.....	112
3.4	ENSAMBLE DE MODELOS.....	112
3.5	SELECCIÓN DEL PUNTO DE CORTE PARA EL CRIBADO....	115
3.6	CRIBADO VIRTUAL.....	115
3.7	DOMINIO DE APLICACIÓN.....	117
3.8	MÉTODOS DIRECTOS. MODELADO ESTRUCTURAL DE LAS FABPs Y GENERACIÓN DE MODELOS DE DOCKING.....	118
3.8.1	MODELADO ESTRUCTURAL DE LAS FABPs.....	118
3.8.1.1	COMPILADO DE LAS SECUENCIAS PARA CADA ISOFORMA.....	119
3.8.1.2	GENERACIÓN DE LAS ESTRUCTURAS TRIDIMENSIONALES.....	119
3.8.1.3	GENERACIÓN DE ENSAMBLES CONFORMACIONALES POR DINÁMICA MOLECULAR.....	120
3.8.1.4	SELECCIÓN DE CONFORMACIONES.....	121
3.8.1.5	OBTENCIÓN DE SITIOS DE HIDRATACIÓN POR DINÁMICA MOLECULAR.....	123
3.8.2	GENERACIÓN DE MODELOS DE DOCKING.....	124
3.8.2.1	COMPILADO DE LA BASE DE DATOS.....	127
3.8.2.2	PREPARACIÓN DE LOS LIGANDOS.....	128
3.8.2.3	PREPARACIÓN DE LOS RECEPTORES.....	128
3.8.3	VALIDACIÓN DE LOS MODELOS DE DOCKING.....	129
3.8.3.1	VALIDACIÓN DE LA FUNCIÓN DE PUNTUACIÓN... ..	129
3.8.3.2	SELECCIÓN DEL PUNTO DE CORTE PARA EL CRIBADO.....	129
3.8.3.3	ANÁLISIS DE LOS SITIOS DE UNIÓN PARA CADA UNA DE LAS ISOFORMAS.....	130
3.8.3.4	VALIDACIÓN DE LA PREDICCIÓN DE LA POSE EXPERIMENTAL.....	130
3.8.4	CRIBADO VIRTUAL.....	131
3.9	ENSAYOS EXPERIMENTALES.....	131
3.9.1	EXPRESIÓN Y PURIFICACIÓN DE LAS FABPs RECOMBINANTES.....	131
3.9.2	ENSAYOS DE DESPLAZAMIENTO MONITOREADOS POR FLUORESCENCIA.....	134

3.9.3 AJUSTE ESTADÍSTICO Y DETERMINACIÓN DE CONSTANTES DE DISOCIACIÓN.....	135
3.9.4 ENSAYOS DE DICROÍSMO CIRCULAR.....	137
REFERENCIAS.....	139
CAPÍTULO 4: RESULTADOS.....	151
4.1 COMPILADO DEL CONJUNTO DE DATOS.....	151
4.2 PARTICIÓN DEL CONJUNTO DE DATOS.....	152
4.3 GENERACIÓN Y VALIDACIÓN DE MODELOS.....	153
4.4 SELECCIÓN DEL PUNTO DE CORTE PARA EL CRIBADO....	158
4.5 CRIBADO VIRTUAL.....	162
4.6 MODELADO ESTRUCTURAL DE LAS FABPs.....	163
4.7 GENERACIÓN Y SELECCIÓN DE CONFORMACIONES POR DINÁMICA MOLECULAR, PCA Y K-MEANS.....	164
4.8 OBTENCIÓN DE SITIOS DE HIDRATACIÓN POR DINÁMICA MOLECULAR.....	166
4.9 VALIDACIÓN DE LA FUNCIÓN DE PUNTUACIÓN.....	167
4.10 SELECCIÓN DEL VALOR DE CORTE PARA EL CRIBADO..	171
4.11 CRIBADO VIRTUAL.....	174
4.12 VALIDACIÓN DE LA POSE EXPERIMENTAL.....	177
4.13 ENSAYOS DE DESPLAZAMIENTO MONITOREADOS POR FLUORESCENCIA.....	184
4.14 ENSAYOS DE DICROÍSMO CIRCULAR.....	200
REFERENCIAS.....	205
CAPÍTULO 5: CONCLUSIONES.....	208
5.1 CONCLUSIONES GENERALES.....	208
5.2 CONCLUSIÓN FINAL.....	213
REFERENCIAS.....	214
MATERIAL SUPLEMENTARIO.....	217
MATERIAL SUPLEMENTARIO 3.....	217
MATERIAL SUPLEMENTARIO 4.....	229
REFERENCIAS.....	247

ACRÓNIMOS

%BC	Porcentaje de buenas clasificaciones
Å	Angstroms
AD Bias	<i>Autodock Bias</i>
AD GPU	Autodock-GPU
AD Hyd	Autodock hidratado (<i>Hydrated Autodock</i>)
ADN	Ácido desoxirribonucleico
AFABP	Proteína ligadora de ácidos grasos del adipocito (<i>Adipocyte fatty acid binding protein</i>)
ANS	Ácido 8-anilino naftaleno-1-sulfónico
ARN	Ácido ribonucleico
ARNi	Ácido ribonucleico de interferencia
ATP	Adenosín trifosfato
AUCROC	Área bajo la curva Característica Operativa del Receptor (<i>Area under the Receiver Operating Characteristics curve</i>)
BEDROC	Discriminación mejorada de la Característica Operativa del Receptor basada en una distribución de Boltzmann (<i>Boltzmann-Enhanced Discrimination of Receiver Operating Characteristics</i>)
BFGS	Algoritmo de Broyden-Fletcher-Goldfarb-Shanno
CASP	Evaluación crítica de las técnicas para la predicción estructural de proteínas (<i>Critical Assessment of Techniques for Protein Structure Prediction</i>)
CD	Dicroísmo circular (<i>circular dichroism</i>)
CDS	Secuencia codificante (<i>Coding sequence</i>)
Charge	Carga formal
CPU	Unidad de procesamiento central (<i>Central processor unit</i>)
CV	Cribado virtual
CVJ	Cribado virtual jerárquico
CVP	Cribado virtual paralelo
CYP450s	Sistema de citocromos P450
DA	Dominio de aplicación
DAVID	Base de datos para anotación, visualización y descubrimiento integrado (<i>Database for Annotation, Visualization and Integrated</i>

	<i>Discovery)</i>
DBL	Descubrimiento basado en la literatura
DM	Dinámica molecular
DO	Densidad óptica
DRH	Repositorio del reposicionamiento de fármacos (<i>Drug Repurposing Hub</i>)
DUDE-Z	Directorio Mejorado de Señuelos Útiles (<i>Enhanced Directory of Useful Decoys</i>)
DvD	Base de datos de fármaco versus enfermedad (<i>Drug versus disease database</i>)
E	Energía
EC50	Concentración efectiva 50
ECFP4	Huella dactilar molecular de conectividad extendida 4 (<i>Extended Connectivity Fingerprint 4</i>)
ED50	Dosis efectiva 50
EgFABP	Proteína ligadora de ácidos grasos de <i>Echinococcus granulosus</i> (<i>Echinococcus granulosus fatty acid binding protein</i>)
EmFABP	Proteína ligadora de ácidos grasos de <i>Echinococcus multilocularis</i> (<i>Echinococcus multilocularis fatty acid binding protein</i>)
EmuDdi1	Proteína ADN daño-inducible 1 de <i>E.multilocularis</i> (<i>DNA damage-inducible 1 protein</i>)
ENL	Eritemas nudosos leprosos
ETDs	Enfermedades tropicales desatendidas
FABP(s)	Proteína(s) ligadora(s) de ácidos grasos (<i>Fatty acid binding protein(s)</i>)
far-UV CD	Dicroísmo circular en el UV lejano (<i>far-UV circular dichroism</i>)
FDA	Administración de Alimentos y Medicamentos de los Estados Unidos (<i>Food and Drug Administration</i>)
FDs	Funciones discriminantes
fMCS	Fracción de la máxima subestructura común
FN	Falsos negativos
FP	Falsos positivos
GDT	Prueba de distancia global (<i>Global distance test</i>)
GPU	Unidad de procesamiento gráfico (<i>Graphic processor unit</i>)
GR	Glutación reductasa

HDL	Lipoproteínas de alta densidad (<i>High density lipoproteins</i>)
HTS	Cribado de Alto Rendimiento (<i>High Throughput Screening</i>)
IC50	Concentración inhibitoria 50
IPTG	Isopropil- β -D-1-tiogalactopiranosido
iRaPCA	Agrupamiento iterativo basado en Subespacios Aleatorios y Análisis de Componentes Principales (<i>iterative Random subspace Principal Component Analysis clustering</i>)
IUPAC	Unión Internacional de Química Pura y Aplicada (<i>International Union of Pure and Applied Chemistry</i>)
Kapp	Constante aparente de disociación
Ki	Constante de inhibición
LB	Medio <i>Luria-Bertani</i>
LDDT	Prueba de la diferencia de distancia local (<i>Local Distance Difference Test</i>)
LDL	Lipoproteínas de baja densidad (<i>Low density lipoproteins</i>)
LGA	Algoritmo genético lamarckiano (<i>Lamarckian genetic algorithm</i>)
LHS	Lipasa hormona sensible
LINCS	Biblioteca de Firmas Celulares Integradas Basadas en Redes (<i>Library of Integrated Network-Based Cellular Signatures</i>)
Log P	Logaritmo del coeficiente de partición octanol-agua
LUDe	Herramienta de generación de señuelos útiles del LIDeB (<i>LIDeB's useful decoys</i>)
MANTRA	Modo de Acción mediante Análisis de Red (<i>Mode of Action by NeTwoRk Analysis</i>)
MCS	Máxima subestructura común
MIN	Operador mínimo
mLogP	Logaritmo del coeficiente de partición octanol-agua estimado por Moriguchi
MSA	Alineamiento múltiple de secuencias (<i>Multiple sequence alignment</i>)
MW	Peso molecular
N	Número de partículas
near-UV CD	Dicroísmo circular en el UV cercano (<i>near-UV circular dichroism</i>)
nHAcc	Número de átomos aceptores de hidrógeno
nHDon	Número de átomos donores de hidrógeno

NPT	Ensamble isotérmico-isobárico
NR	Fracción no retenida
nRotB	Número de enlaces rotables
NVE	Ensamble microcanónico
NVT	Ensamble canónico
OMS	Organización Mundial de la Salud
PBS	Buffer fosfatos
PC	Computadora personal
PCA	Análisis de Componentes Principales (<i>Principal Component Analysis</i>)
PDB	Banco de Datos de Proteínas (<i>Protein Data Bank</i>)
PPV	Valor Predictivo Positivo
PROD	Operador producto
PROM	Operador promedio
QSAR	Relación Cuantitativa Estructura-Actividad (<i>Quantitative Structure-Activity Relationship</i>)
RAM	Memoria de acceso aleatorio (<i>Random access memory</i>)
RANK	Operador promedio del ranking
RMN	Resonancia magnética nuclear
RMSD	Desviación cuadrática media (<i>Root mean square deviation</i>)
ROC	Característica Operativa del Receptor (<i>Receiver Operating Characteristics curve</i>)
SDS-PAGE	Gel de poliacrilamida para electroforesis con dodecil sulfato de sodio (<i>Sodium dodecyl sulfate polyacrylamide gel electrophoresis</i>)
Se	Sensibilidad
SIL	Métrica Silhouette
SMILES	Sistema simplificado de entrada de estructuras químicas en una línea (<i>simplified molecular-input line-entry system</i>)
Sp	Especificidad
SPR	Resonancia Plasmónica de Superficie (<i>Surface Plasmon Resonance</i>)
SREBP	Proteína de unión a elementos reguladores de esteroides (<i>Sterol response elements binding protein</i>)
TGR	Tiorredoxina glutatión reductasa
TrxR	Tiorredoxina reductasa
TsFABP	Proteína ligadora de ácidos grasos de <i>Taenia solium</i> (<i>Taenia solium</i>)

fatty acid binding protein)

TTD	Base de datos de dianas terapéuticas (<i>Therapeutic Target Database</i>)
V	Volumen
VC	Volúmenes de columna
Vina	Autodock Vina
VMD	Visualizador de Dinámica Molecular (<i>Visual Molecular Dynamics</i>)
VN	Verdaderos negativos
VOT	Operador promedio del voto
VP	Verdaderos positivos
W	Fracción lavado
WFP	Probabilidad de encontrar agua (<i>Water finding probability</i>)
WS	Sitios de hidratación (<i>Water sites</i>)
Ya	Prevalencia de activos

ESTADÍAS, PUBLICACIONES Y CONGRESOS

Durante el desarrollo de esta tesis doctoral se han realizado 9 presentaciones a congresos y 8 publicaciones científicas con referato. Cabe destacar que el poster titulado “*Druggability assessment algorithm based on Composition, Transition and Distribution descriptors and publicly available predictive tools*” permitió, por un lado, la aplicación a una beca de asistencia al congreso “*Target Validation using Genomics and Informatics conference*”, organizado por la *Wellcome Connecting Science*, del Reino Unido en el año 2022, como así también la posibilidad de presentarlo de manera oral en el congreso “*XI CAB2C Congreso Argentino de Bioinformática y Biología Computacional*”, organizado por la Sociedad Argentina de Bioinformática en el año 2021.

Por otro lado, durante el año 2022 se realizaron dos estadías de tres meses cada una en Europa, en el marco del consorcio **REFRACT** (acrónimo de *Repeat protein Function, Refinement, Annotation and Classification of Topologies*) financiado por dos becas Marie Curie de la Unión Europea (H2020 EU GA N 823886). En los primeros tres meses de estancia fui recibido en el laboratorio del Dr. Pau Bernadó, ubicado en el *Centre de Biologie Structurale (CBS)*, sito en Montpellier, Francia. Allí me incorporé a un proyecto relacionado con la caracterización estructural del factor de transcripción **PHOX2B** mediante resonancia magnética nuclear y dinámica molecular, en el cual sigo en colaboración activa al día de hoy. En los tres meses sucesivos de estancia en Europa tuve la oportunidad de ser recibido en el laboratorio de la Dra. María Anisimova, en el *Institute of Computational Life Sciences (ICLS-ZHAW)*, ubicado en el pueblo de Wädenswil, Suiza. En esta segunda estadía, llevé adelante el estudio de la interacción entre la proteína *Huntingtin* y su partner *Huntingtin associated protein-1* mediante análisis de coevolución.

A continuación se detalla la lista de las publicaciones en las que he formado parte durante el período de esta tesis doctoral:

ARTÍCULOS ORIGINALES

- Ruatta, S. M., Prada Gori, D. N., Fló Díaz, M., Lorenzelli, F., Perelmuter, K., Alberca, L. N., Bellera, C. L., Medeiros, A., López, G. V., Ingold, M., Porcal, W., Dibello, E., Ihnatenko, I., Kunick, C., Incerti, M., Luzardo, M., Colobbio, M., Ramos, J. C., Manta, E., ... Comini, M. A. (2023). Garbage in, garbage out: How reliable training data improved a virtual screening approach against SARS-CoV-2 MPro. *Frontiers in Pharmacology*, 14, 1193282. <https://doi.org/10.3389/fphar.2023.1193282>

- Prada Gori, D. N., Alberca, L. N., Rodriguez, S., Alice, J. I., Llanos, M. A., Bellera, C. L., & Talevi, A. (2022). LIDeB Tools: A Latin American resource of freely available, open-source cheminformatics apps. *Artificial Intelligence in the Life Sciences*, 2, 100049. <https://doi.org/10.1016/j.aillsi.2022.100049>
- BÉlgamo, J. A., Alberca, L. N., Pórfido, J. L., Romero, F. N. C., Rodriguez, S., Talevi, A., Córscico, B., & Franchini, G. R. (2020). Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets. *Journal of Computer-Aided Molecular Design*, 34(12), 1275–1288. <https://doi.org/10.1007/s10822-020-00352-8>

CAPÍTULO DE LIBRO

- Rodríguez, S., Alice, J. I., Bellera, C. L., & Talevi, A. (2022). Structure-Based Binding Pocket Detection and Druggability Assessment. In M. T. Scotti & C. L. Bellera (Eds.), *Drug Target Selection and Validation* (Vol. 1, pp. 83–97). Springer International Publishing. https://doi.org/10.1007/978-3-030-95895-4_5

ARTÍCULOS DE REVISIÓN

- Rivero Berti, I., Gantner, M. E., Rodriguez, S., Islán, G. A., Fávaro, W. J., Talevi, A., Castro, G. R., & Durán, N. (2023). Potential biocide roles of violacein. *Frontiers in Nanotechnology*, 5, 1186386. <https://doi.org/10.3389/fnano.2023.1186386>
- Lombardo, J. F., Pórfido, J. L., Sisti, M. S., Giorello, A. N., Rodríguez, S., Córscico, B., & Franchini, G. R. (2022). Function of lipid binding proteins of parasitic helminths: Still a long road. *Parasitology Research*, 121(4), 1117–1129. <https://doi.org/10.1007/s00436-022-07463-1>
- Llanos, M. A., Gantner, M. E., Rodriguez, S., Alberca, L. N., Bellera, C. L., Talevi, A., & Gavernet, L. (2021). Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: The Main Protease (Mpro) Case Study. *Journal of Chemical Information and Modeling*, 61(8), 3758–3770. <https://doi.org/10.1021/acs.jcim.1c00404>

- Bellera, C. L., Llanos, M., Gantner, M. E., Rodriguez, S., Gavernet, L., Comini, M., & Talevi, A. (2021). Can drug repurposing strategies be the solution to the COVID-19 crisis? *Expert Opinion on Drug Discovery*, 16(6), 605–612. <https://doi.org/10.1080/17460441.2021.1863943>

OBJETIVO GENERAL

Desarrollar nuevos compuestos terapéuticos aplicables a las enfermedades zoonóticas equinocosis, teniasis y cisticercosis. A tales efectos, en esta tesis doctoral se plantea el desarrollo de modelos computacionales capaces de identificar potenciales inhibidores de las proteínas ligadoras de ácidos grasos (del inglés *Fatty Acid Binding Proteins*, **FABPs**). Se pretende aplicar dichos modelos en el cribado virtual (**CV**) de grandes bases de datos públicas de compuestos químicos. Se busca conseguir nuevos potenciales agentes terapéuticos capaces de ser selectivos, y con actividad durante los distintos estadios de estas enfermedades.

OBJETIVOS ESPECÍFICOS

1. Predicción de inhibidores específicos dirigidos contra FABPs de cestodos. Estrategias basadas en las estructuras de ligandos.

A partir de las estructuras moleculares de ligandos conocidos de las **FABPs**, se generarán, mediante métodos de aprendizaje automático supervisado, funciones discriminantes capaces de identificar compuestos provenientes de bases de datos públicas que tengan la capacidad de unirse a las **FABPs** de cestodos. En base a los resultados, se procederá a la evaluación experimental de los compuestos candidatos frente a isoformas recombinantes de las **FABPs**.

2. Modelado de las FABPs de cestodos con estructura terciaria desconocida.

El análisis de la información obtenida a través del secuenciamiento de los genomas de cestodos, ha permitido la identificación de varios genes que codifican para **FABPs** en los genomas de *E. granulosus*, *E. multilocularis* y *T. solium*. Miembros de nuestro grupo han identificado las secuencias codificantes de cinco 5 isoformas de **FABP** en cada uno de las especies mencionadas.

Hasta el momento se han clonado y secuenciado varias en nuestro laboratorio, pero se dispone de la estructura a nivel atómico de una sola: **EgFABP1** de *E. granulosus*. **EgFABP1** es la única **FABP** de *Echinococcus spp.* cristalizada hasta la actualidad. La disponibilidad de esta estructura, facilita la búsqueda de ligandos que puedan interferir con la función de la proteína mediante métodos basados en la estructura. La búsqueda de agentes farmacológicos que modifiquen la función de **FABPs** específicas de parásitos puede contribuir al control de vías de señalización mediada por lípidos, respuestas inflamatorias y regulación metabólica de fundamental

importancia para el desarrollo y supervivencia del parásito. Por lo antes expuesto, proponemos llevar a cabo el análisis estructural de las isoformas de **FABPs** cuya estructura terciaria es aún desconocida.

3. Predicción de inhibidores específicos dirigidos contra FABPs de cestodos. Estrategias basadas en las estructuras de las FABPs.

Se llevará a cabo la búsqueda de ligandos específicos para **FABPs** de parásitos. Para tales fines, se utilizará la técnica de dinámica molecular (**DM**) para la optimización de los modelos obtenidos en el objetivo 2, y por otro lado, dichos modelos optimizados serán empleados para la clasificación de los compuestos provenientes de bases de datos públicas en activos o inactivos según en función de su capacidad predicha de unión a los modelos de las **FABPs** mediante *docking* molecular. En base a los resultados, se procederá a la evaluación experimental de los compuestos candidatos frente a isoformas recombinantes de las **FABPs**.

CAPÍTULO 1: INTRODUCCIÓN

1.1 ENFERMEDADES TROPICALES DESATENDIDAS

Las enfermedades tropicales desatendidas (**ETDs**) son un grupo de enfermedades infecciosas que afectan a más de mil millones de personas en el planeta, principalmente en países y/o regiones de ingresos bajos y medios (*World Health Organization, Neglected Tropical Diseases, 2023*). Son descritas como desatendidas debido a las limitadas políticas de salud pública dirigidas a su prevención, diagnóstico y tratamiento, como así también a la histórica falta de inversión desde el sector privado debido a que afectan principalmente a poblaciones de bajos recursos y marginadas (*World Health Organization, Neglected Tropical Diseases, 2023*), con escasa expectativa de retorno de inversión (Sbaraglini et al., 2016).

La Organización Mundial de la Salud (**OMS**) lista un conjunto de 20 enfermedades consideradas como desatendidas, las cuales exhiben una serie de características en común:

1. Afectan principalmente a personas viviendo en condiciones de pobreza y/o en zonas rurales o poco pobladas, donde el acceso a centros de salud no es directo.
2. En relación con el punto anterior, constituyen un obstáculo para el motor económico de los países en vías de desarrollo, debido a la dificultad que tienen estas personas de insertarse en el mercado laboral. Hotez y colaboradores (Hotez et al., 2007) desarrollan este concepto en términos de los años de vida ajustados por discapacidad, cuyo valor para las **ETDs** en conjunto es cercano al de enfermedades que representan problemas sanitarios significativos en el mundo entero, como por ejemplo las enfermedades cardiovasculares. Se estima que las **ETDs** generan pérdidas del orden de los mil millones de dólares estadounidenses por año (Remme et al., 2006; Bleakley, 2007; Ramaiah et al., 2000; Frick et al., 2003).
3. Pueden generar estigmatización debido a las deformaciones y/o discapacidades causadas por la enfermedad.
4. El diagnóstico y tratamientos disponibles son limitados e ineficaces.
5. Tienen un alto porcentaje de mortalidad.
6. Podrían ser controladas, prevenidas y erradicadas si se tomaran las medidas sanitarias pertinentes.

Con el propósito de abordar las **ETDs**, la **OMS** propone como medidas concretas la entrega de medicamentos a personas afectadas, el control de los vectores infecciosos, implementación de políticas de salud pública veterinaria, provisión de acceso a agua potable y condiciones de higiene adecuadas (*World Health*

Organization, *Neglected Tropical Diseases*, 2023). La **Figura 1.1** muestra un esquema de las **ETDs** categorizadas por el tipo de agente etiológico.

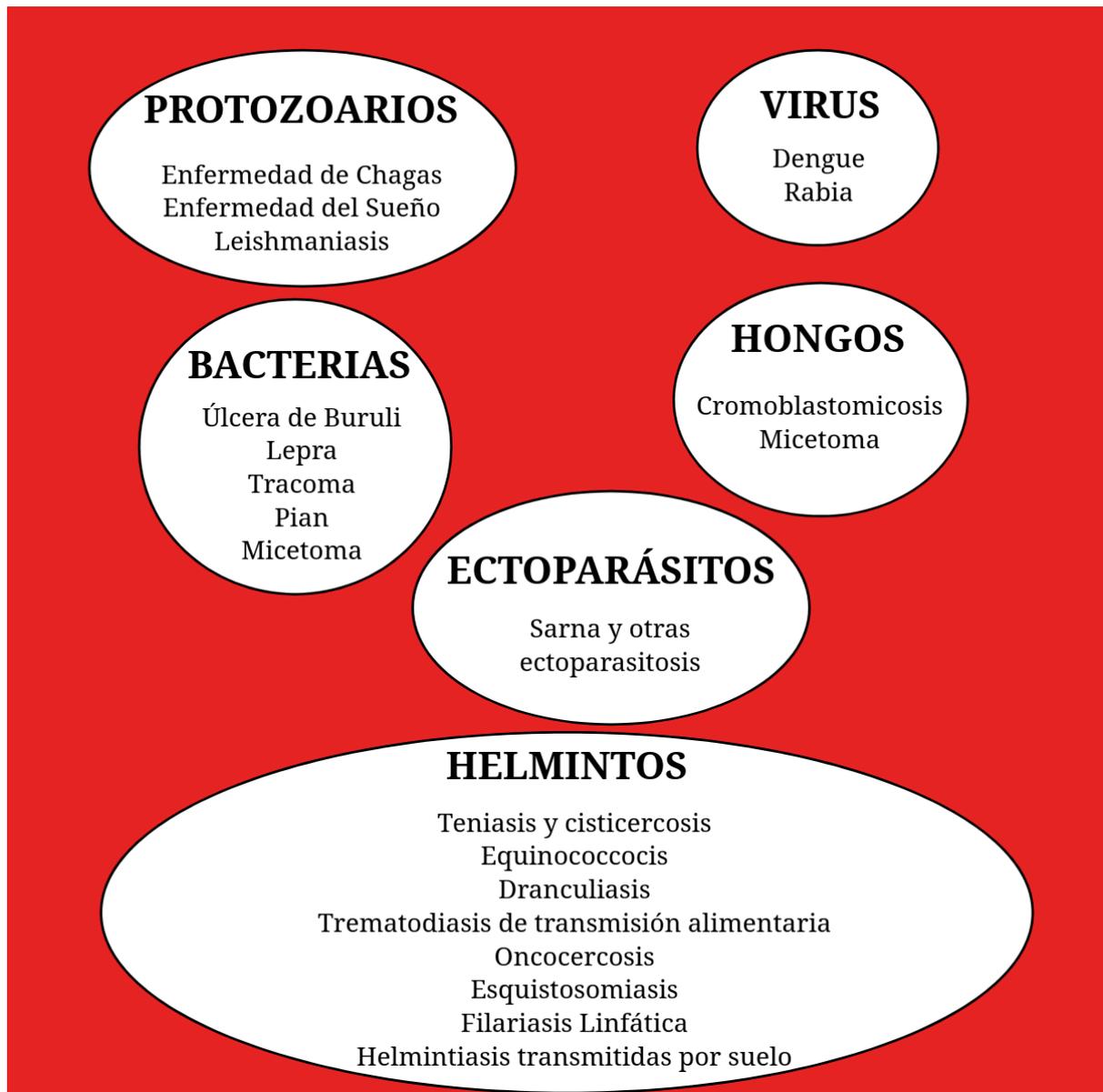


Figura 1.1: ETDs según la OMS.

Esta tesis doctoral procura realizar un aporte al conocimiento actual de nuevas terapias para la equinococcosis, teniasis y cisticercosis, tres **ETDs** de gran prevalencia en nuestro país y en Latinoamérica.

1.2 EQUINOCOCOSIS, TENIASIS Y CISTICERCOSIS

1.2.1 GENERALIDADES

La equinococosis es una parasitosis causada por helmintos del género *Echinococcus*. Dicha parasitosis se manifiesta de manera clínica principalmente en dos formas (CDC, *Echinococcosis-Resources for Health Professionals*, 2020):

1. Equinococosis quística o hidatidosis, producto de la infección por *Echinococcus granulosus*. Los quistes se alojan en hígado y pulmón como localizaciones primarias.
2. Equinococosis alveolar, causada por la infección de *Echinococcus multilocularis*. En este caso, la localización primaria de los quistes es en hígado.

La transmisión de los parásitos de este género se produce a través de la ingesta de huevos del parásito presentes en alimentos y aguas contaminadas. El periodo asintomático de incubación para ambas patologías puede durar muchos años, hasta que aparezcan los primeros signos clínicos. La hidatidosis presenta una tasa de incidencia anual de aproximadamente 50 por 100.000 personas, con una prevalencia de entre 5-10% en zonas rurales de Argentina, Perú, África Oriental, Asia Central y China, mientras que la equinococosis alveolar se restringe a países del hemisferio norte. El tratamiento de la equinococosis, en sus dos formas clínicas, combina la intervención quirúrgica complementada con el uso de benzimidazoles, aunque en muchos casos estos no sean bien tolerados (Gottsten et al., 2015) o su eficacia varíe dependiendo del tipo, tamaño y localización del quiste infeccioso (CDC, *Echinococcosis-Resources for Health Professionals*, 2020).

Por otra parte, la teniasis es una parasitosis intestinal causada por parásitos del género *Taenia*. Si bien el ser humano puede ser infectado por los parásitos *Taenia saginata*, *Taenia asiatica* o *Taenia solium*, los dos primeros carecen de relevancia clínica, mientras que el último puede generar un cuadro clínico severo, conocido como neurocisticercosis. La transmisión de la teniasis por *T. solium* se produce por la ingesta de carne infestada de cerdo cruda o poco cocinada, como así también a través de huevos del parásito presentes en agua y/o alimentos contaminados. Las larvas, denominadas cisticercos, pueden desarrollarse en diversos tejidos, entre ellos el sistema nervioso central, lo cual provoca el cuadro de neurocisticercosis que constituye la causa global prevenible de epilepsia más frecuente (*World Health Organization, Taeniasis/Cysticercosis*, 2023).

La cisticercosis es endémica en áreas rurales de África, Asia y América Latina, donde además de afectar a la salud de la población también produce pérdidas en la ganadería porcina. El tratamiento actual de la neurocisticercosis consiste en la terapia con Praziquantel y/o Albendazol, junto con terapia de soporte con corticoesteroides y/o fármacos anticonvulsivantes y cirugía. Como en el caso de la equinococosis, la efectividad del tratamiento varía de acuerdo al número, tamaño, localización y estadio de los cisticercos (*World Health Organization, Taeniasis/Cysticercosis*, 2023).

1.2.2 *E. GRANULOSUS*, *E. MULTIOCCULARIS* Y *T. SOLIUM*

La equinococosis, como ya se ha mencionado con anterioridad, es causada por *E. granulosus* y *E. multilocularis*, quienes al igual que *T. solium* forman parte del grupo de los helmintos planos. En la **Figura 1.2.2**, se muestra la clasificación taxonómica para cada uno de ellos.

<i>Echinococcus</i> spp.	<i>Taenia solium</i>
Reino: <i>Animalia</i> Filo: <i>Platyhelminthes</i> Subfilo: <i>Rhabditophora</i> Superclase: <i>Neodermata</i> Clase: <i>Cestoda</i> Orden: <i>Cyclophyllidea</i> Familia: <i>Taeniidae</i> Género: <i>Echinococcus</i> Especie: <i>E. granulosus</i> / <i>E. multilocularis</i>	Reino: <i>Animalia</i> Filo: <i>Platyhelminthes</i> Subfilo: <i>Rhabditophora</i> Superclase: <i>Neodermata</i> Clase: <i>Cestoda</i> Orden: <i>Cyclophyllidea</i> Familia: <i>Taeniidae</i> Género: <i>Taenia</i> Especie: <i>T. solium</i>

Figura 1.2.2. Taxonomía de *Echinococcus* spp. y *Taenia solium*.

1.2.3 CICLO DE VIDA DE LOS PARÁSITOS

Los ciclos de vida de *E. granulosus* y *E. multilocularis* se caracterizan por involucrar hospedadores definitivos e intermediarios, en donde el hombre hace las veces de hospedador intermediario circunstancial.

La forma adulta de *E. granulosus*, cuyo tamaño oscila entre los 2-7 mm de largo, reside en el intestino delgado del hospedador definitivo (usualmente perros domésticos) (**Figura 1.2.3.1a**), desde donde se liberan huevos embrionados e infectivos a través de la materia fecal (**Figura 1.2.3.1b**). Dichos huevos pueden ser ingeridos por el hospedador intermediario propio del ciclo (por ejemplo, las ovejas) o por el hombre. Cualquiera sea el caso, los huevos eclosionan en el intestino delgado del hospedador intermediario dando lugar a un nuevo estadio larval, conocido como oncosfera, que tiene la capacidad de adherirse a la pared intestinal, penetrarla y migrar hacia a diversos órganos, principalmente hígado y pulmón, vía circulación sistémica (**Figura 1.2.3.1c**). Es allí donde la oncosfera se desarrolla hacia el conocido quiste hidatídico, una estructura de forma aproximadamente esférica/ovoidal que con el paso del tiempo va produciendo protoescólices y quistes hijos en su interior (**Figura 1.2.3.1d**). La infección hacia el hospedador definitivo se

da por el consumo de carne cruda o mal cocida contaminada con quistes hidatídicos, provenientes del hospedador intermediario (**Figura 1.2.3.1e**). Una vez ingeridos, los protoescólices eclosionan del quiste y se anclan a la pared intestinal del hospedador definitivo, en donde se desarrollan hacia el estadio adulto en un período que oscila entre los 32 a 80 días, cerrando así el ciclo biológico del parásito (**Figura 1.2.3.1a**) (CDC, *Echinococcosis*, 2019).

Los seres humanos, como ya se mencionó, pueden desarrollar el quiste hidatídico en diversos órganos. En el caso de ruptura del quiste, los protoescólices liberados a circulación sistémica pueden crear quistes secundarios en otros sitios del organismo, lo que se conoce como equinococosis secundaria. Este cuadro representa entre 1-8% de los casos, siendo una complicación grave de la enfermedad (Castro-Boix et al., 2008). El ciclo de vida de *E. granulosus* está esquematizado en la **Figura 1.2.3.1**.

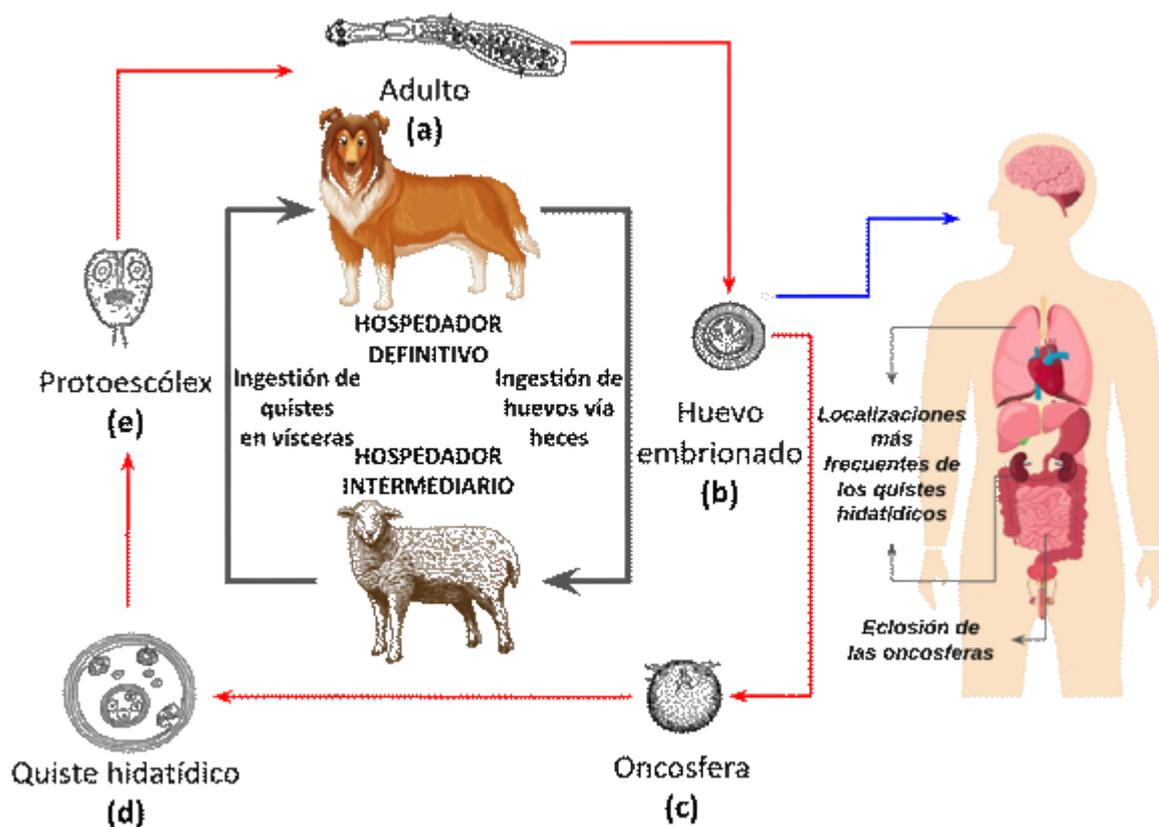


Figura 1.2.3.1. Ciclo de vida de *E. granulosus*. Figura adaptada de CDC, *Echinococcosis*, 2019.

El ciclo de vida de *E. multilocularis* exhibe muchas similitudes con el de *E. granulosus*, aunque con algunas diferencias. El gusano adulto de *E. multilocularis* tiene un tamaño de entre 1,2 a 4,5 mm de longitud, menor en comparación al estadio homólogo de *E. granulosus*. Otra diferencia radica en los hospedadores

intermediarios y definitivos, los cuales en este caso suelen ser principalmente roedores (campañoles y lemmis) y cánidos silvestres (lobos, zorros), respectivamente. Por otro lado, la localización del quiste de *E. multilocularis* se da de forma prácticamente exclusiva en el hígado. Finalmente, el estadio de oncosfera de *E. multilocularis* tiene la capacidad de generar quistes multilobulados, a diferencia de los quistes uniloculares producidos por *E. granulosus*. El ciclo de vida de *E. multilocularis* es esquematizado en la **Figura 1.2.3.2** (CDC, Echinococcosis, 2019).

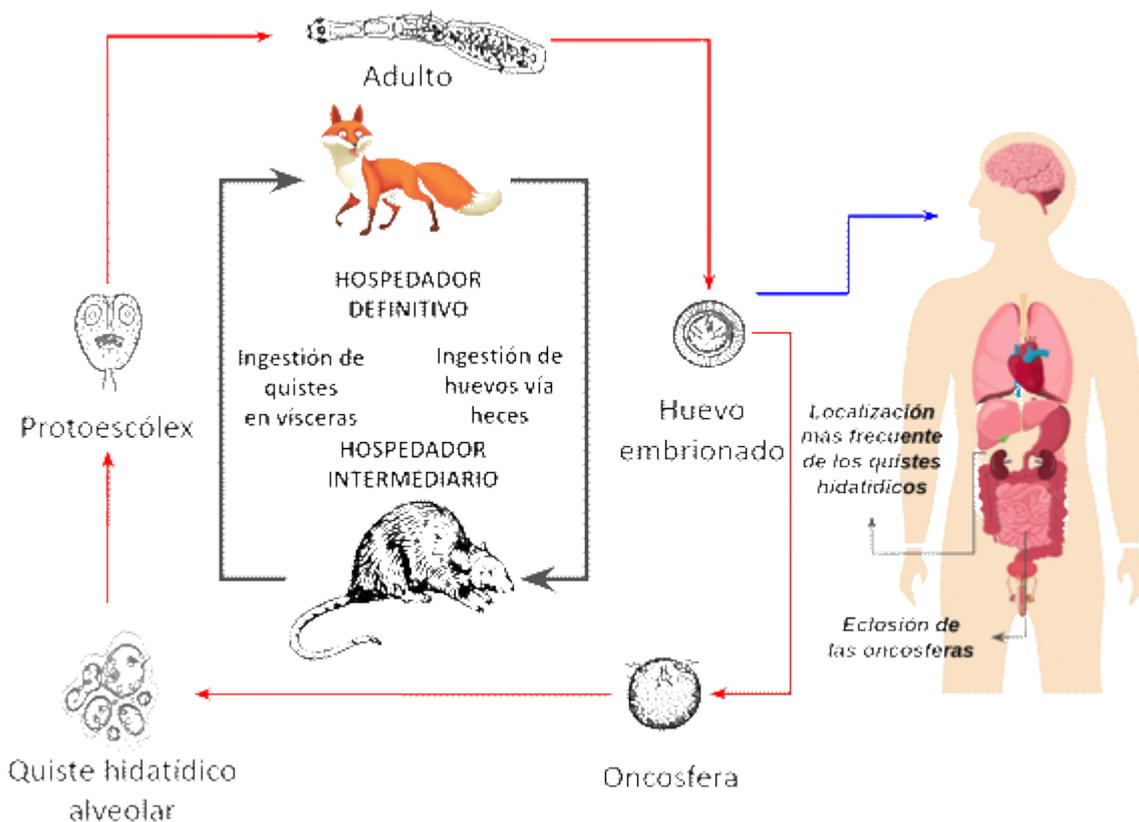


Figura 1.2.3.2. Ciclo de vida de *E. multilocularis*. Figura adaptada de CDC, *Echinococcosis*, 2019.

El ciclo de vida de *T. solium* se muestra en la **Figura 1.2.3.3**. La infección es causada tanto en humanos como en cerdos por la ingesta de huevos embrionados provenientes de agua y/o alimentos contaminados (**Figura 1.2.3.3a**). En el caso de los humanos, la infección también se suele dar por vía fecal-oral causada por malas condiciones de higiene. Una vez ingeridos los huevos (**Figura 1.2.3.3b** y **Figura 1.2.3.3g**, respectivamente), éstos eclosionan en el intestino delgado del hospedador en forma de oncosferas. Las oncosferas se anclan al intestino delgado, invaden la pared intestinal y viajan por circulación sistémica (**Figura 1.2.3.3c** y **Figura 1.2.3.3h**) para alcanzar múltiples tejidos en los cuales pueden diferenciarse en cisticercos en

un plazo de entre 60 a 70 días (**Figura 1.2.3.3d** y **Figura 1.2.3.3i**). En el caso de que los cisticercos alcancen el sistema nervioso central, se desarrollará el cuadro clínico de neurocisticercosis. Por otra parte, si la infección se da por consumo de carne porcina mal cocida, la misma se considera una teniasis, ya que en este caso el estadio infeccioso es el cisticerco y no los huevos embrionados (**Figura 1.2.3.3e**). Consecuentemente, los cisticercos se desarrollan hacia parásitos adultos en el intestino delgado, los cuales son capaces de liberar huevos embrionados a través de la materia fecal y así cerrar el ciclo biológico del parásito (CDC, *Cysticercosis*, 2019).

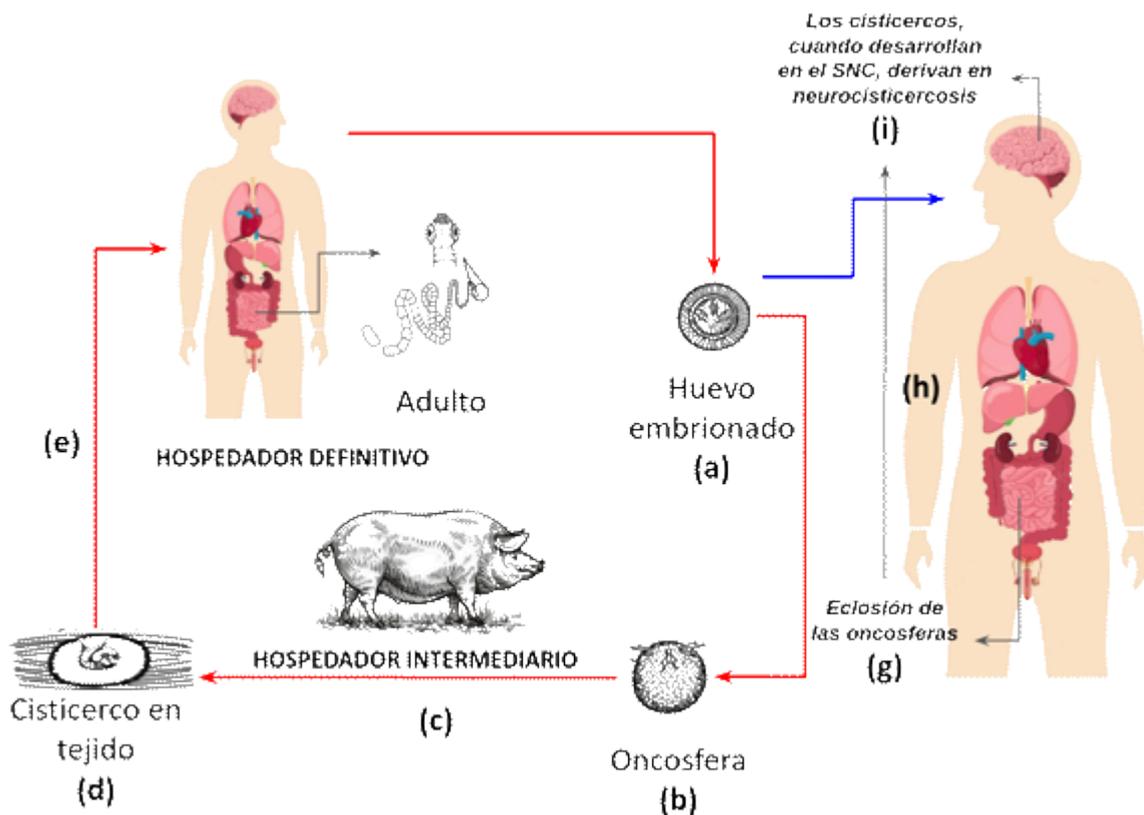


Figura 1.2.3.3. Ciclo de vida de *T. solium*. Figura adaptada de CDC, *Cysticercosis*, 2019.

1.2.4 EPIDEMIOLOGÍA Y DISTRIBUCIÓN

E. granulosus se distribuye geográficamente en todos los continentes, aunque las regiones de mayor prevalencia a nivel mundial son Rusia, China, la región mediterránea asiática, el nordeste de África, Australia y Sudamérica.

En Sudamérica, la hidatidosis representa un problema de salud serio en Argentina, Bolivia, Perú, Brasil y Uruguay. En nuestro país, tiene una prevalencia mayor en las zonas rurales, principalmente en sitios de cría de ganado ovino y caprino. Se estima

que en aproximadamente el 30% de nuestro territorio el ciclo de *E. granulosus* se encuentra activo (Ministerio de Salud de la Nación Argentina, 2012).

Según el Centro Panamericano de Fiebre Aftosa y Salud Pública Veterinaria, en el año 2018 fueron reportados 5687 casos de hidatidosis humana en Argentina, Chile, Perú y Uruguay, siendo nuestro país el segundo en cantidad de casos detectados (553 casos, 9,7% del total). La **Figura 1.2.4.1** presenta la distribución de tales casos por provincia durante el 2018 (*Pan American Health Organization, 2020*).

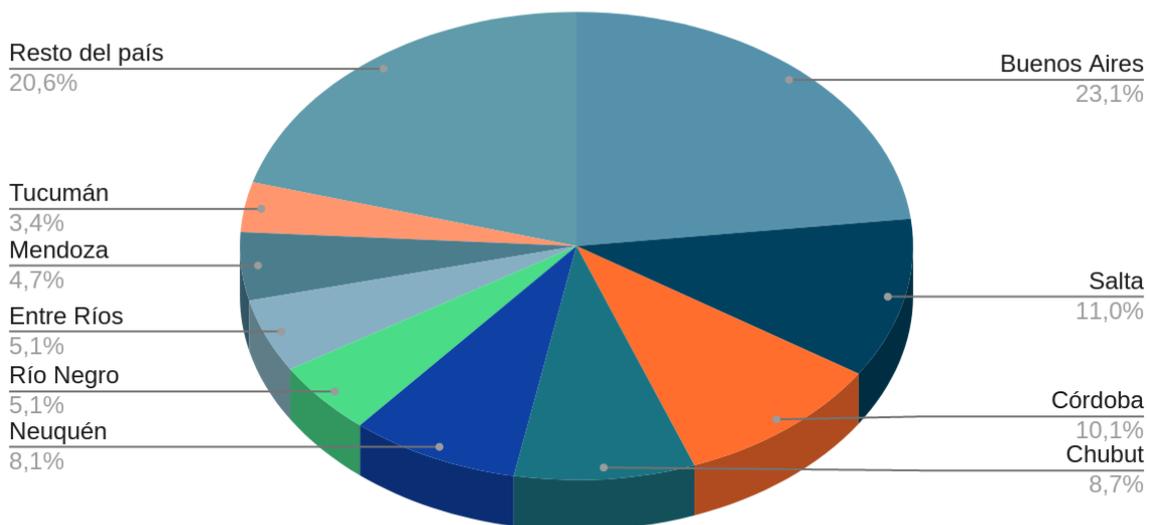


Figura 1.2.4.1. Distribución de casos de hidatidosis en Argentina en el año 2018.

Por otro lado, *E. multilocularis* es un parásito que se encuentra principalmente en el hemisferio norte y cuyo ciclo de vida se desarrolla en la fauna salvaje (Deplazes et al., 2017). En Europa, la prevalencia en zorro rojo oscila entre 0-10% en los distintos países, mientras que en zonas altamente endémicas la prevalencia llega al 50%. En Asia, el foco endémico más importante es en China (Craig et al., 1992), con una tasa del 5% de contagio en algunas poblaciones (Budke et al., 2004). Por otra parte, en Asia central se han reportado numerosos casos de *E. multilocularis* en la población de perros (Budke et al., 2005), lo cual es un indicador de potencial riesgo para los humanos.

Finalmente, *T. solium* ha sido un problema serio en América Latina durante décadas (Bruno et al., 2013), siendo también endémica en el sudeste asiático (Rajshekhar et al., 2003). La **OMS** estima que alrededor de 50 millones de personas en el mundo tienen epilepsia, de las cuales el 30% serían a causa de neurocisticercosis en países endémicos (Ndimubanzi et al., 2010).

En nuestro país, no existen datos oficiales para esta enfermedad, pero se han comunicado 17 casos solamente en el Hospital de Clínicas de la Ciudad de Buenos Aires (Bergroth et al., 2007) y se reportaron 7 casos en el Servicio de Neurología del Hospital HIGA de La Plata durante los años 2002-2005 (Barreiro et al., 2006). En el Hospital Señor del Milagro, Salta, se han registrado 12 casos autóctonos de neurocisticercosis entre los años 2000 y 2003 (Decima et al., 2004). Las crecientes migraciones desde áreas endémicas, principalmente Bolivia y Perú, hacen que esta zoonosis deba ser tenida en cuenta en nuestro medio.

1.2.5 MANIFESTACIONES CLÍNICAS

La infección por *E. granulosus* produce quistes hidatídicos con mayor frecuencia en el hígado y los pulmones y menos frecuentemente en otros órganos como el sistema nervioso central, riñones o huesos. La enfermedad tiene un período asintomático de incubación que puede durar muchos años hasta dar lugar a los primeros signos clínicos. Cuando la localización de los quistes es hepática, el paciente puede sentir dolor abdominal, náuseas y vómitos, mientras que cuando la localización es pulmonar, los signos clínicos son tos crónica, dolor torácico y disnea (World Health Organization, *Echinococcosis*, 2023). La ruptura de quistes dentro del huésped puede causar fiebre, urticaria, eosinofilia y shock anafiláctico, además de la diseminación del quiste hacia otros órganos (CDC, *Echinococcosis*, 2019).

Por otro lado, la infección por *E. multilocularis* se caracteriza por un período de incubación carente de síntomas de entre 5 y 15 años, con desarrollo hacia una lesión tipo tumoral en el hígado. La sintomatología va desde pérdida de peso, dolor abdominal, malestar general e insuficiencia hepática (World Health Organization, *Echinococcosis*, 2023). No es frecuente la diseminación de la larva hacia pulmones, bazo o cerebro, aunque infecciones no tratadas tienen una alta tasa de mortalidad (CDC, *Echinococcosis*, 2019).

Por último, en la cisticercosis, el período de incubación es variable entre los diferentes casos y los síntomas pueden no aparecer por años. En ciertas regiones endémicas, como en Asia, la infección puede manifestarse como nódulos palpables a nivel subcutáneo, mientras que la neurocisticercosis puede causar cefalea crónica, ceguera, convulsiones, hidrocefalia, meningitis y/o demencia. La sintomatología de la neurocisticercosis dependerá del número, tamaño, estadio y localización de los cisticercos, como así también de la respuesta inmune del huésped (World Health Organization, *Taeniasis/Cysticercosis*, 2023). La neurocisticercosis puede ser clasificada como parenquimatosa (afecta a la sustancia gris cerebral) o extraparenquimatosa (afecta a las meninges, ventrículos, cisternas basales, espacio subaracnoideo o médula espinal), lo cual es de gran importancia en relación con el pronóstico del paciente. La enfermedad parenquimatosa con un número bajo de quistes tiene excelente pronóstico a largo plazo, en contraposición a la enfermedad

parenquimatosa con más de 50 quistes o la extraparenquimatosa (CDC, *Cysticercosis-Resources for Health Professionals*, 2022).

En el caso de la teniasis, la infección puede cursar con dolor abdominal, náuseas, y/o diarrea debido a los parásitos adultos presentes en el intestino delgado, aproximadamente ocho semanas después de la ingesta de los cisticercos. Estos síntomas pueden perdurar hasta la muerte del parásito por tratamiento farmacológico, mientras que en casos no tratados la sintomatología puede durar entre 2 a 3 años (*World Health Organization, Taeniasis/Cysticercosis*, 2023).

1.2.6 MEDICAMENTOS DISPONIBLES

En la actualidad, existen dos fármacos, el Albendazol y Praziquantel, que se usan en la terapéutica tanto de la equinococosis (quística y alveolar) (Velasco-Tirado et al., 2018) como de la neurocisticercosis. Por otra parte, se utiliza la Niclosamida para tratar la teniasis al igual que el Albendazol y Praziquantel (*World Health Organization, Taeniasis/Cysticercosis*, 2023).

El Albendazol (**Figura 1.2.6**) es un compuesto orgánico del grupo de los ésteres del Ácido 2-benzimidazol carbámico. Estos compuestos son heteropoliciclos aromáticos que contienen un grupo éster del ácido carbámico, el cual presenta un enlace N al carbono 2 del esqueleto de benzimidazol. El mecanismo de acción descrito para el Albendazol es su efecto inhibitorio en la polimerización de la tubulina, resultando en una pérdida de los microtúbulos citoplasmáticos parasitarios. El Albendazol se une al sitio de unión de la colchicina de la tubulina inhibiendo así la polimerización de los microtúbulos. De esta manera, la ausencia de microtúbulos afectan la captación de glucosa por parte del parásito, llevando a la muerte del mismo (Jacob et al., 2022).

Para el tratamiento de la equinococosis, el Ministerio de Salud de la Nación (Ministerio de Salud de la Nación Argentina, 2012) indica el tratamiento con Albendazol en pacientes asintomáticos, como profilaxis prequirúrgica o en pacientes sintomáticos con cirugía contraindicada. Cualquiera sea el caso, la dosis recomendada es de entre 10-15 mg/kg/día, en dos tomas diarias cada 12 horas luego de una comida con contenido graso para mejorar la absorción del fármaco. El tratamiento estándar tiene una duración de 3 meses, pudiéndose extender hasta 6, salvo que haya intolerancia o que los análisis de laboratorio presenten valores alterados en las enzimas hepáticas.

El efecto adverso más frecuente es la intolerancia gastrointestinal, como también la elevación de las transaminasas hepáticas. La hepatotoxicidad de los benzimidazoles ya fue reportada por Brunetti et al. (2010) y Kern (2010), como así también que la eficacia de la terapéutica está relacionada al tipo, tamaño y localización del quiste (CDC, *Echinococcosis-Resources for Health Professionals*, 2020).

El Praziquantel (**Figura 1.2.6**) es un compuesto perteneciente a la familia de las tetrahidroisoquinolinas. Praziquantel aumenta la permeabilidad del tegumento de los parásitos, logrando un incremento del calcio intra tegumental, lo que lleva a rápidas contracciones y parálisis de la musculatura del verme (Thomas & Timson, 2020).

La administración conjunta de Albendazol y Praziquantel ha mostrado una mayor actividad escolicida en relación a la monoterapia con Albendazol (Velasco-Tirado et al., 2018). La terapia consiste en la administración de 40 mg/kg de Praziquantel en diferentes regímenes para cada paciente (diario, semanal, quincenal o mensual) con dosis estándar de Albendazol durante 2 o 3 meses, sin embargo es necesaria una mayor cantidad de estudios para determinar el costo-beneficio de la terapia combinada en comparación con la monoterapia con Albendazol. Aparentemente, la terapia combinada incrementa los niveles séricos del Albendazol sulfóxido, el metabolito activo del Albendazol, en comparación con pacientes que reciben únicamente la monoterapia con Albendazol (Velasco-Tirado et al., 2018).

En el caso de la neurocisticercosis, la combinación de Praziquantel y Albendazol ha demostrado ser más efectiva que la monoterapia con Albendazol (García et al., 2014). En ese estudio, 124 pacientes con neurocisticercosis intraparenquimal fueron asignados al azar en tres grupos, que recibieron durante 10 días una terapia combinada de Albendazol y Praziquantel (15 mg/kg/día y 50 mg/kg/día, respectivamente), dosis estándar de Albendazol (15 mg/kg/día) o una dosis incrementada de Albendazol (22,5 mg/kg/día). El 64% de los pacientes que recibieron la terapia combinada presentaron remisión completa de los quistes cerebrales, contra el 37% con la terapia estándar de Albendazol y el 53% con la dosis incrementada de Albendazol.

La Niclosamida (**Figura 1.2.6**), por su parte, pertenece al grupo de las benzanilidas. Éstos son compuestos de tipo aromático que contienen un grupo anilida en el cual el grupo carbaxamida está sustituido por un anillo de benceno, siendo su estructura genérica $RNC(=O)R'$, donde R,R' son anillos de benceno (García et al., 2014). Esta familia de compuestos actúan como agentes desacoplantes de la fosforilación oxidativa del parásito adulto, provocando así su inanición en el intestino del hospedero (Weinbach et al., 1969; Williamson et al., 1965; Frayha et al., 1997; Swan, 1999). En seres humanos, la Niclosamida tiene mínima absorción a nivel intestinal, no detectándose sus metabolitos en orina ni en sangre. Se administra por vía oral en presentaciones de 500 mg, y presenta efectos adversos que van de leves a moderados, tales como náuseas, vómito, diarrea y dolor abdominal (García et al., 2014).

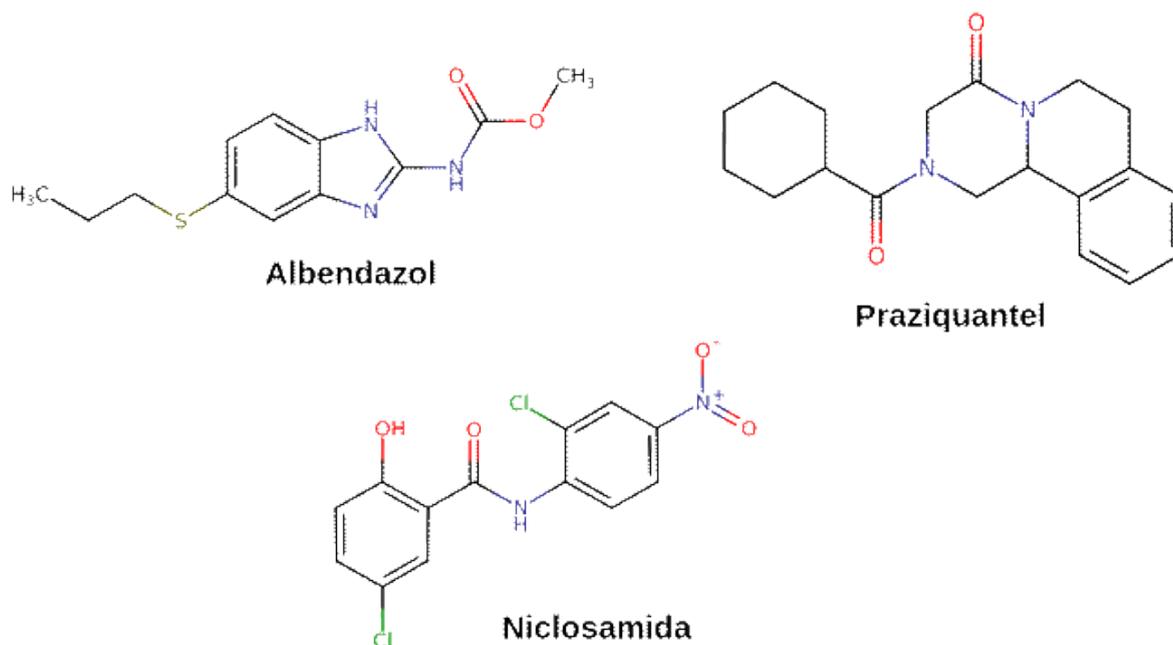


Figura 1.2.6. Estructuras moleculares de los fármacos utilizados para el tratamiento de infecciones por *Echinococcus spp.* y *T. solium*.

Desafortunadamente, se han reportado casos de fármaco-resistencia contra benzimidazoles en veterinaria, vinculados a mutaciones en el gen de β -tubulina del agente infeccioso (Demeler et al., 2013; Geurden et al., 2014). Se han reportado asimismo alelos de resistencia a benzimidazoles en vermes después del tratamiento antihelmíntico, lo que sugiere un fenómeno de resistencia adquirida (Zuccherato et al., 2018; Orr et al., 2019). Por otro lado, se han reportado efectos adversos al Albendazol en la quimioterapia a largo plazo, entre ellos disfunción hepática, leucopenia y caída del cabello (Meilinger et al., 2013). Como se puede percibir, el desarrollo de nuevas terapéuticas contra este tipo de enfermedades es clave para mejorar las perspectivas de los pacientes (Partridge et al., 2020).

1.2.7 AVANCES HACIA NUEVOS MEDICAMENTOS

Dada la naturaleza de las **ETDs**, en términos de su distribución y prevalencia, el sector farmacéutico no encuentra incentivos económicos para la investigación y desarrollo de nuevas terapias para combatirlas (Sbaraglini et al., 2016). A su vez, en el supuesto de que un nuevo fármaco surgiera como potencial agente terapéutico, realizar estudios clínicos aleatorizados o de cohorte sería una tarea compleja de llevar a cabo, dado el alto grado de subdiagnóstico en **ETDs** como la equinococosis (Xu et al., 2022). Por dichas razones, el reposicionamiento de fármacos es una idea atractiva para el tratamiento de las **ETDs** (Polamreddy & Gattu, 2019).

El reposicionamiento de fármacos es un enfoque farmacológico que busca encontrar nuevos usos terapéuticos para fármacos ya existentes. Los compuestos candidatos a reposicionamiento suelen ofrecer ciertas ventajas sobre el descubrimiento *de novo*

en términos de los costos asociados al desarrollo de una nueva entidad química (Breckenridge et al., 2019), como por ejemplo contar ya con estudios clínicos desarrollados con anterioridad. El enfoque de reposicionamiento de fármacos será abordado en mayor detalle en el capítulo 2 de la presente tesis.

A continuación, se han rescatado de literatura algunos fármacos candidatos por reposicionamiento para el tratamiento de la equinococosis:

1. **Mefloquina:** Este medicamento antimalárico logró una reducción significativa en el crecimiento parasitario en ratones inoculados por vía intraperitoneal en comparación con el tratamiento estándar con Albendazol, aunque todavía no se ha dilucidado el mecanismo de acción de este fármaco (Lundström-Stadelmann et al., 2020).
2. **Nitazoxanida:** Es utilizada para la infección por parásitos como *Cryptosporidium parvum* y *Giardia lamblia*. Se observó actividad antiparasitaria en experimentos *in vitro* (no se observó crecimiento de quistes en 6 meses de tratamiento con Nitazoxanida y Albendazol) e *in vivo* (no se encontraron metacestodos del quiste hidatídico en ratones BALB/c) (Mai et al., 2011).
3. **Anfotericina B:** Este antifúngico logró inhibir el crecimiento de *Echinococcus spp.* en un modelo *in vitro*. Se agregó Anfotericina B a un cultivo con tejido infectado con metacestodos con 10 semanas de crecimiento, logrando la destrucción de las vesículas a los 8 días del tratamiento. Como desventaja, este fármaco presenta nefrotoxicidad asociada al uso prolongado (Reuter et al., 2003; Xu et al., 2022). Por otra parte, se ha visto que la administración conjunta con Albendazol no provoca la destrucción vesicular, por lo que se recomienda el uso de ambos fármacos de manera secuencial (Reuter et al., 2010).

De la misma manera, a continuación se presentan algunos fármacos candidatos por reposicionamiento para el tratamiento de la teniasis y/o cisticercosis:

1. **Nitazoxanida:** Se reportó actividad contra *T. saginata* en los años '80 (Maisonneuve et al., 1984), siendo un potente agente contra la teniasis, sobre todo en pacientes con resistencia a Praziquantel y/o Niclosamida (Vermund et al., 1986; Lateef et al., 2008). Por otro lado, Lima et al. (2019) observaron cambios morfológicos en *T. crassiceps* mediante ensayos *in vitro*, evidenciando una relación entre el metabolismo gluconeogénico del parásito y el tratamiento con Nitazoxanida.
2. **Tribendimidina:** Es un antihelmíntico de gran espectro con el que se observó una tasa de curación del 67% en 15 pacientes infectados por *Taenia spp.* con la ingesta de una sola dosis oral de 200 mg o 400 mg, dependiendo de la edad de los pacientes (menores y mayores a 15 años respectivamente). Sin embargo, dado el reducido tamaño de la muestra más estudios son requeridos (Steinmann et al., 2008).

3. **Tamoxifeno:** Este fármaco, utilizado para el tratamiento del cáncer de mama, es un modulador de la actividad de estrógenos. Se demostró que el estradiol tiene un rol esencial en la regulación de la reproducción asexual de los cisticercos de *T. crassiceps* en ratón (Terrazas et al., 1994), razón por la cual se decidió ensayar al Tamoxifeno en un modelo *in vivo* de cisticercos. El Tamoxifeno logró reducir la carga parasitaria en un 80% y 50% en ratones hembra y macho, respectivamente, como así también se demostró la pérdida de movilidad y viabilidad de los cisticercos (Vargas-Villavicencio et al., 2007; Del Río-Araiza et al., 2022). Más aún, se reportó que el Tamoxifeno disminuyó el establecimiento de *T. solium* en el intestino de hámsters en aproximadamente el 70%, así como cambios morfológicos en las tenias sobrevivientes (Galileo Escobedo et al., 2013).

1.3 BLANCOS MOLECULARES EN CESTODOS

Como ya se discutió, los fármacos utilizados en la terapéutica de las infecciones causadas por *Echinococcus spp.* y *T. solium* presentan algunos inconvenientes, como por ejemplo efectos adversos o casos de fármaco-resistencia. Por estas razones, la búsqueda de nuevos blancos moleculares y compuestos biológicamente activos es una necesidad imperiosa (Tsai et al., 2013).

La elección de un buen blanco molecular depende del objetivo, aunque, de manera genérica, podrían mencionarse los siguientes requisitos para un fármaco ideal orientado a tratar una enfermedad infecciosa (Gashaw et al., 2012; Talevi et al., 2019):

1. **Esencial:** El parásito debe ser incapaz de sobrevivir cuando el blanco molecular es inhibido farmacológica o genéticamente.
2. **Selectivo:** Debe estar presente en el patógeno y ausente en el huésped, o bien deben observarse diferencias estructurales significativas homólogo en el hospedador, en particular en lo relativo a los sitios de unión a ligando.
3. **Baja resistencia:** Idealmente debería presentar una sola copia del gen que codifica el blanco, no presentar isoformas ni vías de recuperación o *bypass*.
4. **Ser ensayable (en inglés, assayability):** Deben existir ensayos que permitan evaluar la actividad del blanco frente a los candidatos a inhibidor.
5. **Bioquímica favorable:** Se desea que su contenido intracelular sea bajo, una baja tasa de recambio y que la vía inhibida sea un cuello de botella para el desarrollo del patógeno.
6. **Ser drogable (en inglés, druggable):** El blanco molecular debe ser sensible a la inhibición por moléculas tipo fármaco, logrando desencadenar una respuesta farmacológica contra el agente etiológico de la enfermedad.

7. **Conocer su estructura tridimensional:** Puede ser experimental (resonancia magnética nuclear (**RMN**), cristalografía de rayos X, crio-microscopía electrónica) o teórica.

La esencialidad, la selectividad y la drogabilidad de un blanco molecular son requisitos fundamentales para que éste último pueda ser tomado como tal, mientras que las demás características no son estrictamente necesarias, aunque sí recomendadas para que la estrategia farmacológica tenga mayores probabilidades de éxito (Gashaw et al., 2012; Talevi et al., 2019).

A continuación, se describen algunos ejemplos de blancos moleculares propuestos para cestodos descritos en la literatura.

1.3.1 INHIBIDORES DE LA RESPIRACIÓN ANAERÓBICA

Los sistemas biológicos necesitan energía para poder llevar adelante sus funciones vitales. La mayor parte de la energía en las células es producida por la vía de la fosforilación oxidativa, en la cual típicamente interviene la cadena de transporte de electrones con el oxígeno como aceptor final del poder reductor generado en las distintas vías catabólicas.

Los parásitos helmintos, a diferencia de sus hospedadores mamíferos, lograron desarrollar sistemas de respiración anaeróbica en respuesta a la baja tensión de oxígeno presente en el intestino delgado de sus hospedadores (Kita et al., 1997; Tielens et al., 2002).

En el trabajo de Matsumoto et al. (2008), los autores evidencian la existencia del sistema NADH-fumarato reductasa en mitocondrias de *E. multilocularis*. Este sistema es un complejo enzimático que forma parte de una cadena respiratoria capaz de generar ATP en ausencia de oxígeno. En este estudio, ensayan una serie de compuestos derivados de quinazolina, y encuentran actividad *in vitro* contra protoescolices para dos de dichos derivados. Este blanco molecular resulta atractivo para el diseño de fármacos, ya que no se encuentra en mamíferos y es esencial para la vida del parásito. Recientemente, en el trabajo de Enkai y colaboradores (Enkai et al., 2023) se ha propuesto al meroterpenoide ascofuranona como un inhibidor dual de la respiración aeróbica y anaeróbica en *E. multilocularis*, brindando aún más evidencia sobre la importancia de este blanco.

1.3.2 TIORREDOXINA GLUTATIÓN REDUCTASA

El sistema de los citocromos P450 (**CYP450s**) se encuentra ampliamente distribuido, desde las bacterias hasta los mamíferos. Los **CYP450s** son los

encargados de metabolizar un gran abanico de sustancias endógenas tales como esteroides, ácidos grasos y/o prostaglandinas, como así también sustancias exógenas como pueden ser fármacos. Se sabe que la alta actividad de los **CYP450s** está relacionada directamente con la capacidad de eliminación de fármacos y otros xenobióticos mediante metabolismo, y que se trata de uno de los mecanismos responsables de la tolerancia a éstos últimos (Pakharukova et al., 2012). Tsai et al. (2013) mencionan que al igual que los trematodos, los parásitos cestodos tienen un solo gen de **CYP450s**, sugiriendo una limitada habilidad para metabolizar sustancias de naturaleza diversa. A su vez, postula la existencia de la enzima tiorredoxina glutatión reductasa (**TGR**), una de las enzimas que interviene en la homeostasis redox y que combina dos actividades enzimáticas provenientes de dos sistemas diferentes presentes en mamíferos, la glutatión reductasa (**GR**) y la tiorredoxina reductasa (**TrxR**).

En base a la importancia de mantener el balance redox y las diferencias en cuanto a los mecanismos redox entre el parásito trematodo *S. mansoni* y el hospedador, Kuntz et al. (2007) propusieron a **TGR** como blanco molecular para la schistosomiasis. Muy brevemente, en el mencionado trabajo realizaron experimentos de **ARN** de interferencia (**ARNi**) silenciando la expresión del gen **TGR**, como así también probaron varios compuestos con actividad reportada contra sulfuro reductasas en varios organismos. Con respecto a los experimentos de **ARNi**, los investigadores observaron al 92% de los parásitos muertos al 4to día post-tratamiento, en comparación al 5% del experimento control. También encontraron que el fármaco auranofina presentó actividad contra **TGR** en ensayos de inhibición enzimática, en ensayos *in vitro* contra el parásito adulto y en ratones inoculados con *S. mansoni*. Años más tarde, Ross et al. (2012) publicaron un trabajo en el cual ensayan una serie de compuestos contra la **TGR** de *E. granulosus* y *F. hepatica*, como así también contra las larvas respectivas. Como resultado, encontraron compuestos de la familia de los furoxanos y quinoxalinas como compuestos candidatos para ambos parásitos. Toda esta evidencia hace a **TGR** un blanco validado en trematodos y cestodos (Tsai et al., 2013).

1.3.3 PROTEÍNA ADN DAÑO-INDUCIBLE 1 DE *E. MULTILOCULARIS*

La proteína ADN daño-inducible 1 de *E. multilocularis* (del inglés *E. multilocularis DNA damage-inducible 1 protein*, **EmuDdi1**) está reportada como blanco molecular contra las infecciones de dicho cestode (Liu et al., 2022). Proteínas homólogas a **EmuDdi1** en diferentes especies cumplen funciones celulares relacionadas al sistema proteosomal de la ubiquitina y a la inhibición de la secreción de proteínas (Gabriely et al., 2008; Yip et al., 2020).

Ensayos de silenciamiento con **ARNi** sobre **EmuDdi1** mostraron un descenso de la tasa de supervivencia al 60% en protoescolices *in vitro* durante las primeras 72 hs

de cultivo, valor estadísticamente inferior al del grupo control. A su vez, el tratamiento con **ARNi** sobre **EmuDdi1** durante 12 hs provocó un aumento en la secreción proteica, lo cual fue evidenciado en el medio de cultivo donde se encontraban los protoescólices (Liu et al., 2022). Este hecho concuerda con la función inhibitoria sobre la secreción ya reportada anteriormente (Gabriely et al., 2008; Yip et al., 2020). En la misma sintonía, utilizando el antirretroviral Nelfinavir se observó inhibición en la actividad proteolítica de **EmuDdi1** en ensayos de actividad enzimática *in vitro*. Ensayos ulteriores del Nelfinavir en presencia de protoescólices en cultivo arrojaron un valor de concentración efectiva 50 (**EC50**) de 28,84 μM , apreciablemente menor al del Albendazol (68,84 μM). Finalmente, el Nelfinavir fue ensayado en ratones BALB/c tanto inmunocompetentes como inmunodeprimidos, mostrando una disminución en el peso de los quistes, como así también una reducción en el tamaño del bazo, siendo la inflamación esplénica un signo de infección por helmintos (Liu et al., 2022; Cosenza-Contreras et al., 2019).

1.4 PROTEÍNAS DEL METABOLISMO LIPÍDICO

1.4.1 GENERALIDADES

Según la **IUPAC** (en inglés *International Union of Pure and Applied Chemistry*) (Moss et al., 1995), los lípidos se definen como un “conjunto de biomoléculas cuya característica común es la solubilidad en solventes no polares”. Dada la heterogeneidad estructural, los lípidos cumplen diversas funciones a nivel celular como reservorio de energía, señalización y principales componentes estructurales de membranas biológicas. Las distintas categorías de lípidos se muestran en la **Figura 1.4.1**.

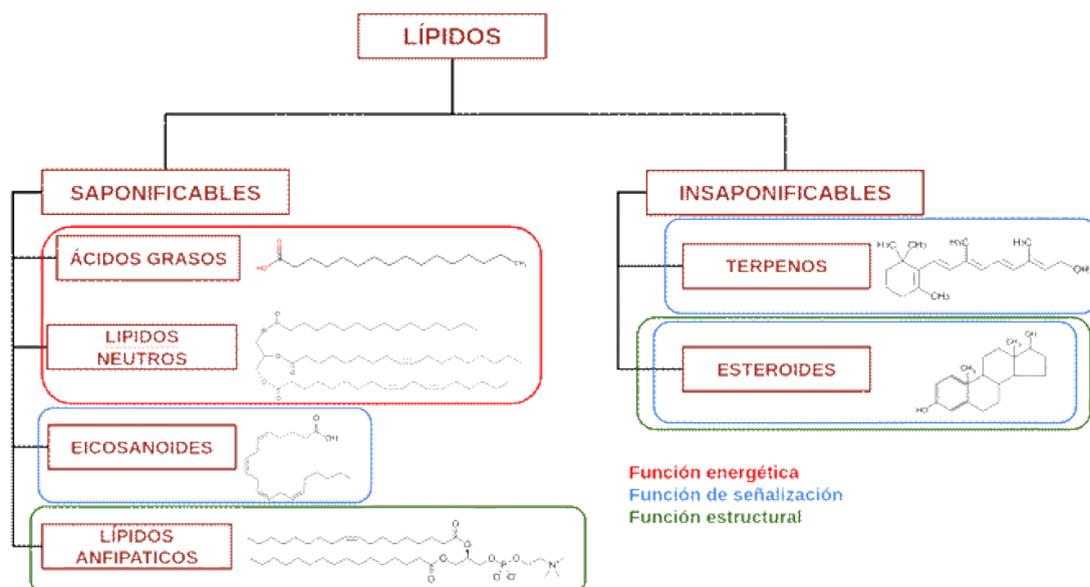


Figura 1.4.1. Clasificación de los distintos tipos de lípidos.

1.4.2 METABOLISMO DE LÍPIDOS

El metabolismo lipídico en humanos es un proceso complejo que involucra la síntesis, el transporte, el almacenamiento y la oxidación de lípidos.

La síntesis de lípidos ocurre principalmente en el hígado y en el tejido adiposo, y es regulada por una serie de enzimas y factores de transcripción. Uno de los principales reguladores de la síntesis de lípidos es la proteína de unión a elementos reguladores de esteroides (**SREBP**, del inglés *sterol response elements binding protein*), que activa la expresión de genes involucrados en la síntesis de ácidos grasos y triglicéridos (Horton et al., 2002).

Los lípidos también se transportan en la sangre en forma de lipoproteínas, que son complejos macromoleculares de lípidos y proteínas. Las lipoproteínas de alta densidad (**HDL**, acrónimo de *high density lipoproteins*) transportan lípidos desde los tejidos periféricos al hígado, mientras que las lipoproteínas de baja densidad (**LDL**, del inglés *low density lipoproteins*) transportan lípidos desde el hígado a los tejidos periféricos (Feingold, 2000).

El almacenamiento de lípidos ocurre en el tejido adiposo, donde los ácidos grasos son esterificados y almacenados como triglicéridos. La lipólisis, que es la hidrólisis de los triglicéridos, es regulada por la lipasa hormona sensible (**LHS**), proporcionando una fuente de ácidos grasos para la oxidación y la producción de energía (Althaher, 2022).

La oxidación de lípidos ocurre en las mitocondrias, donde los ácidos grasos son convertidos en acetil-CoA a través de la beta-oxidación. El acetil-CoA resultante se somete entonces al ciclo de Krebs para la producción de energía en forma de **ATP** (Talley et al., 2023).

Es sabido que los cestodos carecen de la capacidad de sintetizar ácidos grasos y colesterol *de novo*, razón por la cual deben obtener dichas biomoléculas del hospedador (Tsai et al., 2013). Para dar un caso, el análisis del genoma de *E. granulosus* reveló la ausencia tres enzimas clave en el metabolismo de los esteroides, como la hidroximetilglutaril-CoA reductasa, escualeno sintasa y escualeno monooxigenasa (Zheng et al., 2013). La primera de estas enzimas cataliza la conversión del hidroximetilglutaril-CoA a mevalonato, paso limitante de la síntesis de colesterol en mamíferos, mientras que la segunda cataliza la condensación de dos moléculas de farnesil pirofosfato a escualeno. La escualeno reductasa es la enzima que dirige al farnesil pirofosfato hacia la síntesis de esteroides en detrimento de la síntesis de terpenos más complejos. Por último, la escualeno monooxigenasa convierte al escualeno en 2,3 óxido-escualeno, estando esta enzima regulada directamente por los niveles intracelulares de colesterol (Coates et al., 2021). Asimismo, *E. granulosus* carece de varias enzimas para la síntesis de ácidos grasos *de novo* (Zheng et al., 2013), entre ellas la enzima ácido graso

sintasa, la cual se encarga de la síntesis del Ácido palmítico a partir de acetil y malonil-CoA. En la **Figura 1.4.2** se esquematizan ambas vías de síntesis y las principales enzimas ausentes.

Como contrapartida, los parásitos cestodos han desarrollado diversos sistemas para incorporar tales precursores desde el hospedador como por ejemplo transportadores de ácidos grasos, enzimas de elongación de ácidos grasos preexistentes, como así también la expresión de las **FABPs** y la apolipoproteína antígeno B (Tsai et al., 2013). La incorporación de ácidos grasos parece ser clave en los cestodos, dado que tanto las **FABPs** como el antígeno B se encuentran entre los 19 genes más expresados.

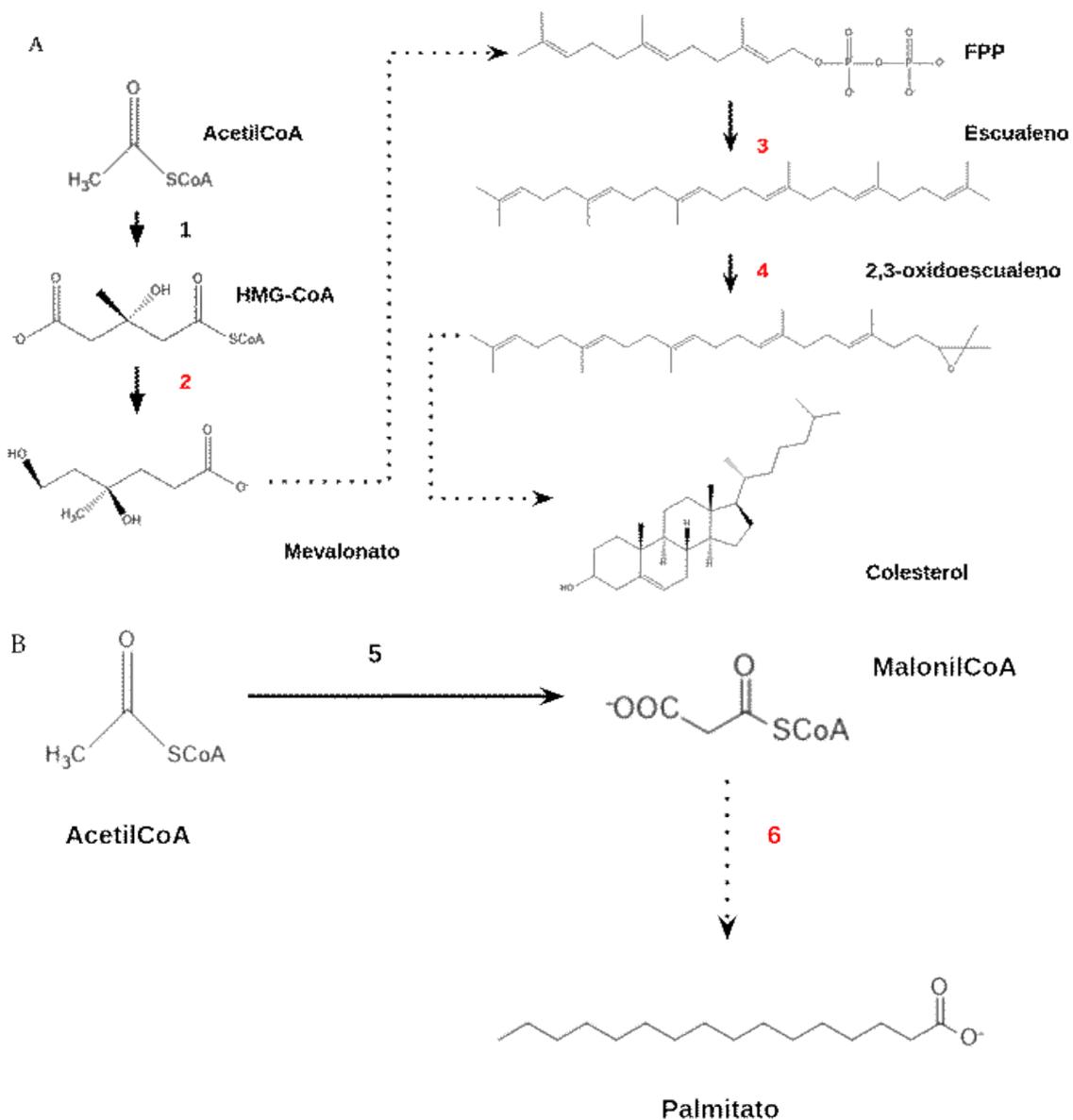


Figura 1.4.2. Esquema simplificado para las vías de síntesis de colesterol (A) y ácidos grasos (B) en mamíferos. Las flechas discontinuas indican varios procesos y/o enzimas involucradas, mientras que los números en color rojo son las enzimas

ausentes en cestodos. HMG-CoA: 3-hidroxi-3-metilglutaril-CoA, FPP: Farnesil pirofosfato. 1: HMG-CoA sintasa, 2: HMG-CoA reductasa, 3: Escualeno sintasa, 4: Escualeno monooxigenasa, 5: AcetilCoA carboxilasa, 6: Ácido graso sintasa

1.4.3 FABPs

Las **FABPs** son pequeñas proteínas intracelulares, con un peso que oscila entre los 14-15 kDa. Como su nombre lo indica, unen principalmente ácidos grasos aunque también pueden ligar diferentes tipos de ligandos hidrofóbicos, entre ellos, esteroides. Las **FABPs** han sido encontradas exclusivamente en el reino animal, con una amplia distribución entre los distintos grupos dentro de este reino. Estas proteínas han sido exhaustivamente estudiadas por estar asociadas a patologías humanas como la obesidad, aterosclerosis y diabetes (Lee et al., 2021), y a su vez, han sido propuestas como posibles blancos farmacológicos en helmintos (Becker et al., 1994; Morphew et al., 2016).

Trabajos previos han mostrado la presencia de seis genes codificantes de las **FABPs** en *Echinococcus spp.* En el caso de *E. multilocularis*, dos de esos genes codifican para la misma proteína, mientras que para *E. granulosus* todas las secuencias codificantes son diferentes (Pórfido et al., 2020). En el caso de *T. solium*, se han identificado secuencias codificantes para **FABPs** en Wormbase (WormBase ParaSite, *Taenia solium*, 2023) usando las seis secuencias codificantes de *E. granulosus* como secuencias de consulta (en inglés *query*). De dicho análisis, se encontraron cinco isoformas para *T. solium*. En el estudio realizado por Pórfido et al. (2012), se realizó una caracterización funcional de **EgFABP1**, una de las isoformas de **FABP** presente en *E. granulosus*, revelando la capacidad de esta proteína de unir y transportar ácidos grasos utilizando diferentes tipos de membranas artificiales. Al día de la escritura de esta tesis doctoral, la isoforma **EgFABP1** es la única **FABP** de cestodos que cuenta con una estructura tridimensional resuelta en el *Protein Data Bank* (PDB:1O8V).

La estructura cristalina de **EgFABP1** presenta el plegamiento tipo barril beta típico de las **FABPs**. La isoforma de **FABP** humana en adipocito (**AFABP**) es la que más parecido estructural guarda con **EgFABP1**, con una identidad de secuencia de aproximadamente el 44%. **EgFABP1** comparte con **AFABP** el motivo P2, el cual consiste en una tríada compuesta por dos argininas y una tirosina, que interacciona con los grupos carboxílicos de los ligandos hidrofóbicos que se unen. En dicha tríada las argininas están separadas por entre sí por veinte residuos mientras que entre la segunda arginina y la tirosina hay uno solo (R...RxY) (Jakobsson et al., 2003).

Desde el punto de vista del descubrimiento de fármacos, se han diseñado con éxito diferentes tipos de inhibidores contra las isoformas de **FABPs** humanas, entre ellas la **AFABP**, consiguiendo en muchos casos la selectividad de isoforma, es decir

lograr la inhibición de una **FABP** de un tejido en particular sin inhibir de forma apreciable a las demás (Cai et al., 2015; Barf et al., 2009; Lehmann et al., 2004). Con este último punto, se podría pensar en el diseño de moléculas inhibitoras con actividad selectiva sobre las **FABPs** de cestodos por sobre las isoformas del huésped.

1.5 ELECCIÓN DEL BLANCO MOLECULAR

En la presente tesis doctoral se emplea un enfoque computacional para la búsqueda de nuevos compuestos químicos con actividad inhibitoria sobre los parásitos *E. granulosus*, *E. multilocularis* y *T. solium*, causantes de las enfermedades equinococosis y cisticercosis. Dado que las **FABPs** han sido propuestas como blanco molecular contra estos cestodos, y al éxito en cuanto al desarrollo de inhibidores isoforma-específico en **FABPs** humanas, se propone enfocar la búsqueda computacional a inhibidores del metabolismo de esta familia de proteínas.

REFERENCIAS

Althaher, A. R. (2022). An Overview of Hormone-Sensitive Lipase (HSL). *The Scientific World Journal*, 2022, 1–9. <https://doi.org/10.1155/2022/1964684>

Barf, T., Lehmann, F., Hammer, K., Haile, S., Axen, E., Medina, C., Uppenberg, J., Svensson, S., Rondahl, L., & Lundbäck, T. (2009). N-Benzyl-indolo carboxylic acids: Design and synthesis of potent and selective adipocyte fatty-acid binding protein (A-FABP) inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 19(6), 1745–1748. <https://doi.org/10.1016/j.bmcl.2009.01.084>

Barreiro MA.; Salazar Salazar CZ.; Zúccolo LA.; Carignano SG.; Joyce C. Congreso de Zoonosis 2006 La Plata, Argentina

Becker, M. M., Kalinna, B. H., Waive, G. J., & McManus, D. P. (1994). Gene cloning, overproduction and purification of a functionally active cytoplasmic fatty acid-binding protein (Sj-FABPc) from the human blood fluke *Schistosoma japonicum*. *Gene*, 148(2), 321–325. [https://doi.org/10.1016/0378-1119\(94\)90706-4](https://doi.org/10.1016/0378-1119(94)90706-4)

Bergroth, B., Contardo, D. M., Mazzei, M. E., Abuchanab, M. E., & Melero, M. J. (2007). Neurocysticercosis. *Medicina*, 67(4), 377.

Bleakley, H. (2007). Disease and Development: Evidence from Hookworm Eradication in the American South. *The Quarterly Journal of Economics*, 122(1), 73–117. <https://doi.org/10.1162/qjec.121.1.73>

Breckenridge, A., & Jacob, R. (2019). Overcoming the legal and regulatory barriers to drug repurposing. *Nature Reviews Drug Discovery*, 18(1), 1–2. <https://doi.org/10.1038/nrd.2018.92>

Brunetti, E., Kern, P., & Vuitton, D. A. (2010). Expert consensus for the diagnosis and treatment of cystic and alveolar echinococcosis in humans. *Acta Tropica*, 114(1), 1–16. <https://doi.org/10.1016/j.actatropica.2009.11.001>

Bruno, E., Bartoloni, A., Zammarchi, L., Strohmeyer, M., Bartalesi, F., Bustos, J. A., Santivañez, S., García, H. H., Nicoletti, A., & the COHEMI Project Study Group. (2013). Epilepsy and Neurocysticercosis in Latin America: A Systematic Review and Meta-analysis. *PLoS Neglected Tropical Diseases*, 7(10), e2480. <https://doi.org/10.1371/journal.pntd.0002480>

Budke, C. M., Campos-Ponce, M., Qian, W., & Torgerson, P. R. (2005). A canine purgation study and risk factor analysis for echinococcosis in a high endemic region of the Tibetan plateau. *Veterinary Parasitology*, 127(1), 43–49. <https://doi.org/10.1016/j.vetpar.2004.08.024>

Budke CM, Jiamin Q, Zinsstag J, Qian W, & Torgerson PR. (2004). Use of disability adjusted life years in the estimation of the disease burden of echinococcosis for a high endemic region of the Tibetan plateau. *1*(71), 56–64.

Cai, H., Liu, Q., Gao, D., Wang, T., Chen, T., Yan, G., Chen, K., Xu, Y., Wang, H., Li, Y., & Zhu, W. (2015). Novel fatty acid binding protein 4 (FABP4) inhibitors: Virtual screening, synthesis and crystal structure determination. *European Journal of Medicinal Chemistry*, 90, 241–250. <https://doi.org/10.1016/j.ejmech.2014.11.020>

Castro-Boix, S., Santamaría-Gómez, M., Grau-Llacuna, S., Nadal-Guinard, A., & Armengol-Carrasco, M. (2008). Rotura intraperitoneal traumática de quiste hidatídico hepático. *Cirugía Española*, 84(1), 45–46. [https://doi.org/10.1016/S0009-739X\(08\)70604-3](https://doi.org/10.1016/S0009-739X(08)70604-3)

CDC- Centers for Disease Control and Prevention. (2019). DPDx—Cysticercosis. <https://www.cdc.gov/dpdx/cysticercosis/index.html>

CDC- Centers for Disease Control and Prevention. (2019). DPDx—Echinococcosis. <https://www.cdc.gov/dpdx/echinococcosis/index.html>

CDC- Centers for Disease Control and Prevention. (2020). Echinococcosis—Resources for Health Professionals. https://www.cdc.gov/parasites/echinococcosis/health_professionals/index.html

CDC- Centers for Disease Control and Prevention. (2022). Cysticercosis. Resources for Health Professionals. <https://www.cdc.gov/parasites/cysticercosis/es/hcp/index.html>

Coates, H. W., Capell-Hattam, I. M., & Brown, A. J. (2021). The mammalian cholesterol synthesis enzyme squalene monooxygenase is proteasomally truncated to a constitutively active form. *Journal of Biological Chemistry*, 296, 100731. <https://doi.org/10.1016/j.jbc.2021.100731>

Cosenza-Contreras, M., Oliveira E Castro, R. A. D., Mattei, B., Campos, J. M., Gonçalves Silva, G., De Paiva, N. C. N., De Oliveira Aguiar-Soares, R. D., Carneiro, C. M., Afonso, L. C. C., & Castro-Borges, W. (2019). The Schistosomiasis SpleenOME: Unveiling the Proteomic Landscape of Splenomegaly Using Label-Free Mass Spectrometry. *Frontiers in Immunology*, 9, 3137. <https://doi.org/10.3389/fimmu.2018.03137>

Craig PS, Deshan L, Macpherson CNL, Dazhong S, Reynolds D, et al. (1992). A large focus of alveolar echinococcosis in central China. *The Lancet*, 340(8823), 826–831. [https://doi.org/10.1016/0140-6736\(92\)92693-A](https://doi.org/10.1016/0140-6736(92)92693-A)

Decima M, Olivo L, López S, Falco A, et al. (2004) VI Jornadas de enfermedades infecciosas emergentes y reemergentes. *Revista Médica de Salta*; 21:75.

Demeler, J., Krüger, N., Krücken, J., Von Der Heyden, V. C., Ramünke, S., Küttler, U., Miltsch, S., López Cepeda, M., Knox, M., Vercruyssen, J., Geldhof, P., Harder, A., & Von Samson-Himmelstjerna, G. (2013). Phylogenetic Characterization of β -Tubulins and Development of Pyrosequencing Assays for Benzimidazole Resistance in Cattle Nematodes. *PLoS ONE*, 8(8), e70212. <https://doi.org/10.1371/journal.pone.0070212>

Del Río-Araiza, V., Hernández-Bello, R., & Morales-Montor, J. (2022). New Uses for Old Drugs and Their Application in Helminthology. In J. Morales-Montor, V. Hugo Del Río-Araiza, & R. Hernández-Bello (Eds.), *Parasitic Helminths and Zoonoses—From Basic to Applied Research*. IntechOpen. <https://doi.org/10.5772/intechopen.106176>

Deplazes, P., Rinaldi, L., Alvarez Rojas, C. A., Torgerson, P. R., Harandi, M. F., Romig, T., Antolova, D., Schurer, J. M., Lahmar, S., Cringoli, G., Magambo, J., Thompson, R. C. A., & Jenkins, E. J. (2017). Global Distribution of Alveolar and Cystic Echinococcosis. In *Advances in Parasitology* (Vol. 95, pp. 315–493). Elsevier. <https://doi.org/10.1016/bs.apar.2016.11.001>

Enkai, S., Kouguchi, H., Inaoka, D. K., Shiba, T., Hidaka, M., Matsuyama, H., Sakura, T., Yagi, K., & Kita, K. (2023). Killing Two Birds with One Stone: Discovery of Dual Inhibitors of Oxygen and Fumarate Respiration in Zoonotic Parasite, *Echinococcus multilocularis*. *Antimicrobial Agents and Chemotherapy*, e01428-22. <https://doi.org/10.1128/aac.01428-22>

Escobedo, G., Palacios-Arreola, M. I., Olivos, A., López-Griego, L., & Morales-Montor, J. (2013). Tamoxifen Treatment in Hamsters Induces Protection during Taeniosis by *Taenia solium*. *BioMed Research International*, 2013, 1–10. <https://doi.org/10.1155/2013/280496>

Feingold, K. R. (2000). Introduction to Lipids and Lipoproteins. In K. R. Feingold, B. Anawalt, M. R. Blackman, A. Boyce, G. Chrousos, E. Corpas, W. W. de Herder, K. Dhatariya, K. Dungan, J. Hofland, S. Kalra, G. Kaltsas, N. Kapoor, C. Koch, P. Kopp, M. Korbonits, C. S. Kovacs, W. Kuohung, B. Laferrère, ... D. P. Wilson (Eds.), *Endotext*. MDText.com, Inc. <http://www.ncbi.nlm.nih.gov/books/NBK305896/>

Frayha, G. J., Smyth, J. D., Gobert, J. G., & Savel, J. (1997). The mechanisms of action of antiprotozoal and anthelmintic drugs in man. *General Pharmacology: The Vascular System*, 28(2), 273–299. [https://doi.org/10.1016/S0306-3623\(96\)00149-8](https://doi.org/10.1016/S0306-3623(96)00149-8)

Frick, K. D., Jacobson, G. A., & Hanson, C. L. (2003). GLOBAL BURDEN OF TRACHOMA AND ECONOMICS OF THE DISEASE. *The American Journal of Tropical Medicine and Hygiene*, 69(5_suppl_1), 1–10. https://doi.org/10.4269/ajtmh.2003.69.5_suppl_1.0690001

Gabriely, G., Kama, R., Gelin-Licht, R., & Gerst, J. E. (2008). Different Domains of the UBL-UBA Ubiquitin Receptor, Ddi1/Vsm1, Are Involved in Its Multiple Cellular Roles. *Molecular Biology of the Cell*, 19(9), 3625–3637. <https://doi.org/10.1091/mbc.e07-05-0462>

Garcia, H. H., Gonzales, I., Lescano, A. G., Bustos, J. A., Zimic, M., Escalante, D., Saavedra, H., Gavidia, M., Rodriguez, L., Najjar, E., Umeres, H., & Pretell, E. J. (2014). Efficacy of combined antiparasitic therapy with praziquantel and albendazole for neurocysticercosis: A double-blind, randomised controlled trial. *The Lancet Infectious Diseases*, 14(8), 687–695. [https://doi.org/10.1016/S1473-3099\(14\)70779-0](https://doi.org/10.1016/S1473-3099(14)70779-0)

Gashaw, I., Ellinghaus, P., Sommer, A., & Asadullah, K. (2012). What makes a good drug target? *Drug Discovery Today*, 17, S24–S30. <https://doi.org/10.1016/j.drudis.2011.12.008>

Geurden, T., Hoste, H., Jacquiet, P., Traversa, D., Sotiraki, S., Frangipane Di Regalbano, A., Tzanidakis, N., Kostopoulou, D., Gaillac, C., Privat, S., Giangaspero, A., Zanardello, C., Noé, L., Vanimisetti, B., & Bartram, D. (2014). Anthelmintic resistance and multidrug resistance in sheep gastro-intestinal nematodes in France, Greece and Italy. *Veterinary Parasitology*, 201(1–2), 59–66. <https://doi.org/10.1016/j.vetpar.2014.01.016>

Gottstein, B., Stojkovic, M., Vuitton, D. A., Millon, L., Marcinkute, A., & Deplazes, P. (2015). Threat of alveolar echinococcosis to public health – a challenge for Europe. *Trends in Parasitology*, 31(9), 407–412. <https://doi.org/10.1016/j.pt.2015.06.001>

Horton, J. D., Goldstein, J. L., & Brown, M. S. (2002). SREBPs: Activators of the complete program of cholesterol and fatty acid synthesis in the liver. *Journal of Clinical Investigation*, 109(9), 1125–1131. <https://doi.org/10.1172/JCI0215593>

Hotez, P. J., Molyneux, D. H., Fenwick, A., Kumaresan, J., Sachs, S. E., Sachs, J. D., & Savioli, L. (2007). Control of Neglected Tropical Diseases. *New England Journal of Medicine*, 357(10), 1018–1027. <https://doi.org/10.1056/NEJMra064142>

Jacob, J., Siraj, M. A., Steel, A., Tan, G., & Jarvi, S. (2022). Evaluation of the mechanism of action of albendazole on adult rat lungworm (*Angiostrongylus cantonensis*). *Experimental Parasitology*, 242, 108355. <https://doi.org/10.1016/j.exppara.2022.108355>

Jakobsson, E., Alvite, G., Bergfors, T., Esteves, A., & Kleywegt, G. J. (2003). The crystal structure of *Echinococcus granulosus* fatty-acid-binding protein 1. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1649(1), 40–50. [https://doi.org/10.1016/S1570-9639\(03\)00151-1](https://doi.org/10.1016/S1570-9639(03)00151-1)

- Kern, P. (2010). Clinical features and treatment of alveolar echinococcosis: Current Opinion in Infectious Diseases, 23(5), 505–512. <https://doi.org/10.1097/QCO.0b013e32833d7516>
- Kita, K. (1997). Cytochromes in the respiratory chain of helminth mitochondria. International Journal for Parasitology, 27(6), 617–630. [https://doi.org/10.1016/S0020-7519\(97\)00016-7](https://doi.org/10.1016/S0020-7519(97)00016-7)
- Kuntz, A. N., Davioud-Charvet, E., Sayed, A. A., Califf, L. L., Dessolin, J., Arnér, E. S. J., & Williams, D. L. (2007). Thioredoxin Glutathione Reductase from *Schistosoma mansoni*: An Essential Parasite Enzyme and a Key Drug Target. PLoS Medicine, 4(6), e206. <https://doi.org/10.1371/journal.pmed.0040206>
- Lateef, M., Zargar, S. A., Khan, A. R., Nazir, M., & Shoukat, A. (2008). Successful treatment of niclosamide- and praziquantel-resistant beef tapeworm infection with nitazoxanide. International Journal of Infectious Diseases, 12(1), 80–82. <https://doi.org/10.1016/j.ijid.2007.04.017>
- Lee, C.-H., Lui, D. T. W., & Lam, K. S. L. (2021). Adipocyte Fatty Acid-Binding Protein, Cardiovascular Diseases and Mortality. Frontiers in Immunology, 12, 589206. <https://doi.org/10.3389/fimmu.2021.589206>
- Lehmann, F., Haile, S., Axen, E., Medina, C., Uppenberg, J., Svensson, S., Lundbäck, T., Rondahl, L., & Barf, T. (2004). Discovery of inhibitors of human adipocyte fatty acid-binding protein, a potential type 2 diabetes target. Bioorganic & Medicinal Chemistry Letters, 14(17), 4445–4448. <https://doi.org/10.1016/j.bmcl.2004.06.057>
- Lima, N. F., Picanço, G. A., Alves, D. D. S. M. M., Silva, L. D., Isac, E., Costa, T. L., Lino Junior, R. D. S., & Vinaud, M. C. (2019). Oral nitazoxanide treatment of experimental neurocysticercosis induces gluconeogenesis in *Taenia crassiceps* cysticerci. *Acta Tropica*, 190, 361–364. <https://doi.org/10.1016/j.actatropica.2018.12.017>
- Liu, Z., Guo, X., Guo, A., Zhang, S., Zou, Y., Wang, Y., Li, X., He, W., Pu, L., Zhang, S., Zeng, Q., Cai, X., & Wang, S. (2022). HIV protease inhibitor nelfinavir is a potent drug candidate against echinococcosis by targeting Ddi1-like protein. *eBioMedicine*, 82, 104177. <https://doi.org/10.1016/j.ebiom.2022.104177>
- Lundström-Stadelmann, B., Rufener, R., & Hemphill, A. (2020). Drug repurposing applied: Activity of the anti-malarial mefloquine against *Echinococcus multilocularis*. International Journal for Parasitology: Drugs and Drug Resistance, 13, 121–129. <https://doi.org/10.1016/j.ijpddr.2020.06.002>
- Mai, H. A., Wang, Y. H., Zhao, J. M., & Bai, L. (2011). Effect of different drugs on *Echinococcus multilocularis* metacestodes in vitro. 6(29), 419-424.

Maisonneuve, H., & Rossignol, J. F. (1984). Nitazoxanide in the Treatment of *Taenia Saginata* and *Hymenolepis nana* Infections. *The American Journal of Tropical Medicine and Hygiene*, 33(3), 511–512. <https://doi.org/10.4269/ajtmh.1984.33.511>

Matsumoto, J., Sakamoto, K., Shinjyo, N., Kido, Y., Yamamoto, N., Yagi, K., Miyoshi, H., Nonaka, N., Katakura, K., Kita, K., & Oku, Y. (2008). Anaerobic NADH-Fumarate Reductase System Is Predominant in the Respiratory Chain of *Echinococcus multilocularis*, Providing a Novel Target for the Chemotherapy of Alveolar Echinococcosis. *Antimicrobial Agents and Chemotherapy*, 52(1), 164–170. <https://doi.org/10.1128/AAC.00378-07>

Meilinger, M., Stoeckl, C., Pollheimer, M., Kern, P., Reisinger, E. C., Seeber, K., Krause, R., Flick, H., & Hoenigl, M. (2013). Progressive alveolar echinococcosis after discontinuation of anthelmintic therapy. *Parasites & Vectors*, 6(1), 287. <https://doi.org/10.1186/1756-3305-6-287>

Ministerio de Salud de la Nación Argentina. (2012). *Hidatidosis, guía médica*. https://bancos.salud.gob.ar/sites/default/files/2018-10/0000000797cnt-2012-03-29_hidatidosis-guia-medica.pdf

Morphew, R. M., Wilkinson, T. J., Mackintosh, N., Jahndel, V., Paterson, S., McVeigh, P., Abbas Abidi, Syed. M., Saifullah, K., Raman, M., Ravikumar, G., LaCourse, J., Maule, A., & Brophy, P. M. (2016). Exploring and Expanding the Fatty-Acid-Binding Protein Superfamily in *Fasciola* Species. *Journal of Proteome Research*, 15(9), 3308–3321. <https://doi.org/10.1021/acs.jproteome.6b00331>

Moss, G. P., Smith, P. A. S., & Tavernier, D. (1995). Glossary of class names of organic compounds and reactivity intermediates based on structure (IUPAC Recommendations 1995). *Pure and Applied Chemistry*, 67(8–9), 1307–1375. <https://doi.org/10.1351/pac199567081307>

Ndimubanzi, P. C., Carabin, H., Budke, C. M., Nguyen, H., Qian, Y.-J., Rainwater, E., Dickey, M., Reynolds, S., & Stoner, J. A. (2010). A Systematic Review of the Frequency of Neurocytotoxicosis with a Focus on People with Epilepsy. *PLoS Neglected Tropical Diseases*, 4(11), e870. <https://doi.org/10.1371/journal.pntd.0000870>

Orr, A. R., Quagraine, J. E., Suwondo, P., George, S., Harrison, L. M., Dornas, F. P., Evans, B., Caccone, A., Humphries, D., Wilson, M. D., & Cappello, M. (2019). Genetic Markers of Benzimidazole Resistance among Human Hookworms (*Necator americanus*) in Kintampo North Municipality, Ghana. *The American Journal of Tropical Medicine and Hygiene*, 100(2), 351–356. <https://doi.org/10.4269/ajtmh.18-0727>

Pakharukova, M. Y., Ershov, N. I., Vorontsova, E. V., Katokhin, A. V., Merkulova, T. I., & Mordvinov, V. A. (2012). Cytochrome P450 in fluke *Opisthorchis felinus*:

Identification and characterization. *Molecular and Biochemical Parasitology*, 181(2), 190–194. <https://doi.org/10.1016/j.molbiopara.2011.11.005>

Pan American Health Organization. (2020). Equinococosis: Informe epidemiológico en la región de América del Sur - 2018, n.4, 2020. In Informe Equinococosis;n.4, mar 2020 [Technical reports]. OPS. <https://iris.paho.org/handle/10665.2/51942>

Partridge, F. A., Forman, R., Bataille, C. J. R., Wynne, G. M., Nick, M., Russell, A. J., Else, K. J., & Sattelle, D. B. (2020). Anthelmintic drug discovery: Target identification, screening methods and the role of open science. *Beilstein Journal of Organic Chemistry*, 16, 1203–1224. <https://doi.org/10.3762/bjoc.16.105>

Polamreddy, P., & Gattu, N. (2019). The drug repurposing landscape from 2012 to 2017: Evolution, challenges, and possible solutions. *Drug Discovery Today*, 24(3), 789–795. <https://doi.org/10.1016/j.drudis.2018.11.022>

Porfido, J. L., Alvite, G., Silva, V., Kennedy, M. W., Esteves, A., & Corsico, B. (2012). Direct Interaction between EgFABP1, a Fatty Acid Binding Protein from *Echinococcus granulosus*, and Phospholipid Membranes. *PLoS Neglected Tropical Diseases*, 6(11), e1893. <https://doi.org/10.1371/journal.pntd.0001893>

Pórfido, J. L., Herz, M., Kiss, F., Kamenetzky, L., Brehm, K., Rosenzvit, M. C., Córscico, B., & Franchini, G. R. (2020). Fatty acid-binding proteins in *Echinococcus* spp.: The family has grown. *Parasitology Research*, 119(4), 1401–1408. <https://doi.org/10.1007/s00436-020-06631-5>

Rajshekhar, V., Joshi, D. D., Doanh, N. Q., Van De, N., & Xiaonong, Z. (2003). *Taenia solium* taeniosis/cysticercosis in Asia: Epidemiology, impact and issues. *Acta Tropica*, 87(1), 53–60. [https://doi.org/10.1016/S0001-706X\(03\)00055-X](https://doi.org/10.1016/S0001-706X(03)00055-X)

Ramaiah, K. D., Das, P. K., Michael, E., & Guyatt, H. L. (2000). The Economic Burden of Lymphatic Filariasis in India. *Parasitology Today*, 16(6), 251–253. [https://doi.org/10.1016/S0169-4758\(00\)01643-4](https://doi.org/10.1016/S0169-4758(00)01643-4)

Remme, J. H. F., Feenstra, P., Lever, P. R., Medici, A. C., Morel, C. M., Noma, M., Ramaiah, K. D., Richards, F., Seketeli, A., Schmunis, G., van Brakel, W. H., & Vassall, A. (2006). Tropical Diseases Targeted for Elimination: Chagas Disease, Lymphatic Filariasis, Onchocerciasis, and Leprosy. In D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, & P. Musgrove (Eds.), *Disease Control Priorities in Developing Countries* (2nd ed.). The International Bank for Reconstruction and Development / The World Bank. <http://www.ncbi.nlm.nih.gov/books/NBK11745/>

Reuter, S., Beisler, T., & Kern, P. (2010). Combined albendazole and amphotericin B against *Echinococcus multilocularis* in vitro. *Acta Tropica*, 115(3), 270–274. <https://doi.org/10.1016/j.actatropica.2010.04.009>

Reuter, S., Merkle, M., Brehm, K., Kern, P., & Manfras, B. (2003). Effect of Amphotericin B on Larval Growth of *Echinococcus multilocularis*. *Antimicrobial Agents and Chemotherapy*, 47(2), 620–625. <https://doi.org/10.1128/AAC.47.2.620-625.2003>

Ross, F., Hernández, P., Porcal, W., López, G. V., Cerecetto, H., González, M., Basika, T., Carmona, C., Fló, M., Maggioli, G., Bonilla, M., Gladyshev, V. N., Boiani, M., & Salinas, G. (2012). Identification of Thioredoxin Glutathione Reductase Inhibitors That Kill Cestode and Trematode Parasites. *PLoS ONE*, 7(4), e35033. <https://doi.org/10.1371/journal.pone.0035033>

Sbaraglini, ML., Cristina Vanrell, M., Leticia Bellera, C., Benaim, G., Carrillo, C., Talevi, A., & Silvia Romano, P. (2016). Neglected Tropical Protozoan Diseases: Drug Repositioning as a Rational Option. *Current Topics in Medicinal Chemistry*, 16(19), 2201–2222. <https://doi.org/10.2174/1568026616666160216154309>

Steinmann, P., Zhou, X.-N., Du, Z.-W., Jiang, J.-Y., Xiao, S.-H., Wu, Z.-X., Zhou, H., & Utzinger, J. (2008). Tribendimidine and Albendazole for Treating Soil-Transmitted Helminths, *Strongyloides stercoralis* and *Taenia* spp.: Open-Label Randomized Trial. *PLoS Neglected Tropical Diseases*, 2(10), e322. <https://doi.org/10.1371/journal.pntd.0000322>

Swan, G. E. (1999). The pharmacology of halogenated salicylanilides and their anthelmintic use in animals: Review article. *Journal of the South African Veterinary Association*, 70(2), 61–70. <https://doi.org/10.4102/jsava.v70i2.756>

Talevi, A., Carrillo, C., & Comini, M. (2019). The Thiol-polyamine Metabolism of *Trypanosoma cruzi*: Molecular Targets and Drug Repurposing Strategies. *Current Medicinal Chemistry*, 26(36), 6614–6635. <https://doi.org/10.2174/0929867325666180926151059>

Talley, J. T., & Mohiuddin, S. S. (2023). *Biochemistry, Fatty Acid Oxidation*. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK556002/>

Terrazas, L. I., Bojalil, R., Govezensky, T., & Larralde, C. (1994). A Role for 17- β -Estradiol in Immunoendocrine Regulation of Murine Cysticercosis (*Taenia crassiceps*). *The Journal of Parasitology*, 80(4), 563. <https://doi.org/10.2307/3283192>

Thomas, C. M., & Timson, D. J. (2020). The Mechanism of Action of Praziquantel: Can New Drugs Exploit Similar Mechanisms? *Current Medicinal Chemistry*, 27(5), 676–696. <https://doi.org/10.2174/0929867325666180926145537>

Tielens, A. G. M., Rotte, C., Van Hellemond, J. J., & Martin, W. (2002). Mitochondria as we don't know them. *Trends in Biochemical Sciences*, 27(11), 564–572. [https://doi.org/10.1016/S0968-0004\(02\)02193-X](https://doi.org/10.1016/S0968-0004(02)02193-X)

Tsai, I. J., Zarowiecki, M., Holroyd, N., Garcarrubio, A., Sanchez-Flores, A., Brooks, K. L., Tracey, A., Bobes, R. J., Fragoso, G., Sciutto, E., Aslett, M., Beasley, H., Bennett, H. M., Cai, J., Camicia, F., Clark, R., Cucher, M., De Silva, N., Day, T. A., ... Berriman, M. (2013). The genomes of four tapeworm species reveal adaptations to parasitism. *Nature*, 496(7443), 57–63. <https://doi.org/10.1038/nature12031>

Vargas-Villavicencio, J. A., Larralde, C., De León-Nava, M. A., Escobedo, G., & Morales-Montor, J. (2007). TAMOXIFEN TREATMENT INDUCES PROTECTION IN MURINE CYSTICERCOSIS. *Journal of Parasitology*, 93(6), 1512–1517. <https://doi.org/10.1645/GE-1191.1>

Velasco-Tirado, V., Alonso-Sardón, M., Lopez-Bernus, A., Romero-Alegria, Á., Burguillo, F. J., Muro, A., Carpio-Pérez, A., Muñoz Bellido, J. L., Pardo-Lledias, J., Cordero, M., & Belhassen-García, M. (2018). Medical treatment of cystic echinococcosis: Systematic review and meta-analysis. *BMC Infectious Diseases*, 18(1), 306. <https://doi.org/10.1186/s12879-018-3201-y>

Vermund, S. H., MacLeod, S., & Goldstein, R. G. (1986). Taeniasis Unresponsive to a Single Dose of Niclosamide: Case Report of Persistent Infection with *Taenia saginata* and a Review of Therapy. *Clinical Infectious Diseases*, 8(3), 423–426. <https://doi.org/10.1093/clinids/8.3.423>

Weinbach, E. C., & Garbus, J. (1969). Mechanism of Action of Reagents that uncouple Oxidative Phosphorylation. *Nature*, 221(5185), 1016–1018. <https://doi.org/10.1038/2211016a0>

Williamson, R. L., & Metcalf, R. L. (1967). Salicylanilides: A New Group of Active Uncouplers of Oxidative Phosphorylation. *Science*, 158(3809), 1694–1695. <https://doi.org/10.1126/science.158.3809.1694>

World Health Organization. (2023). Neglected tropical diseases. <https://www.who.int/news-room/questions-and-answers/item/neglected-tropical-diseases>

World Health Organization. (2023). Echinococcosis. <https://www.who.int/es/news-room/fact-sheets/detail/echinococcosis>

World Health Organization. (2023). Taeniasis/Cysticercosis. <https://www.who.int/news-room/fact-sheets/detail/taeniasis-cysticercosis>

WormBase ParaSite. (2023). *Taenia solium*. https://parasite.wormbase.org/Taenia_solium_prjna170813/Info/Index/

Xu, X., Qian, X., Gao, C., Pang, Y., Zhou, H., Zhu, L., Wang, Z., Pang, M., Wu, D., Yu, W., Kong, F., Shi, D., Guo, Y., Su, X., Hu, W., Yan, J., Feng, X., & Fan, H. (2022). Advances in the pharmacological treatment of hepatic alveolar echinococcosis: From

laboratory to clinic. *Frontiers in Microbiology*, 13, 953846.
<https://doi.org/10.3389/fmicb.2022.953846>

Yip, M. C. J., Bodnar, N. O., & Rapoport, T. A. (2020). Ddi1 is a ubiquitin-dependent protease. *Proceedings of the National Academy of Sciences*, 117(14), 7776–7781.
<https://doi.org/10.1073/pnas.1902298117>

Zheng, H., Zhang, W., Zhang, L., Zhang, Z., Li, J., Lu, G., Zhu, Y., Wang, Y., Huang, Y., Liu, J., Kang, H., Chen, J., Wang, L., Chen, A., Yu, S., Gao, Z., Jin, L., Gu, W., Wang, Z., ... Wang, S. (2013). The genome of the hydatid tapeworm *Echinococcus granulosus*. *Nature Genetics*, 45(10), 1168–1175. <https://doi.org/10.1038/ng.2757>

Zuccherato, L. W., Furtado, L. F., Medeiros, C. D. S., Pinheiro, C. D. S., & Rabelo, É. M. (2018). PCR-RFLP screening of polymorphisms associated with benzimidazole resistance in *Necator americanus* and *Ascaris lumbricoides* from different geographical regions in Brazil. *PLOS Neglected Tropical Diseases*, 12(9), e0006766.
<https://doi.org/10.1371/journal.pntd.0006766>

CAPÍTULO 2: DESCUBRIMIENTO DE FÁRMACOS

2.1 INTRODUCCIÓN

El descubrimiento y desarrollo de fármacos implica la integración de diferentes disciplinas, como la biología, la química, las ciencias farmacéuticas, el modelado matemático y las ciencias de la computación, para lograr principios activos más seguros y eficaces.

El proceso de desarrollo necesario para introducir un nuevo medicamento en el mercado es largo y costoso, y requiere la colaboración de especialistas procedentes de diversos campos disciplinares (Macalino et al., 2015). El costo total varía según el tipo de medicamento, pero siempre es significativo. Por ejemplo, en el año 2020 las principales compañías farmacéuticas a nivel global invirtieron entre el 15-28% de sus ingresos para investigación y desarrollo, lo que se traduce en un costo de entre 4-10 mil millones de dólares estadounidenses para cada nueva entidad química introducida en el mercado (The Investopedia Team, 2022). Además, el tiempo necesario para completar el proceso, desde el descubrimiento hasta el lanzamiento, oscila entre 12 y 15 años (Macalino et al., 2015; Terry & Lesser, 2015). Asimismo, existe un alto riesgo de fracaso, ya que de cada 10.000 compuestos investigados en las fases iniciales del desarrollo sólo uno superará exitosamente los ensayos clínicos y obtendrá la aprobación regulatoria (Cook et al., 2014).

La **Figura 2.1** resume las diferentes etapas que deben atravesarse en el proceso de desarrollo de un nuevo fármaco, para garantizar su seguridad y eficacia y cumplir con los requisitos regulatorios (Mohs & Greig, 2017).

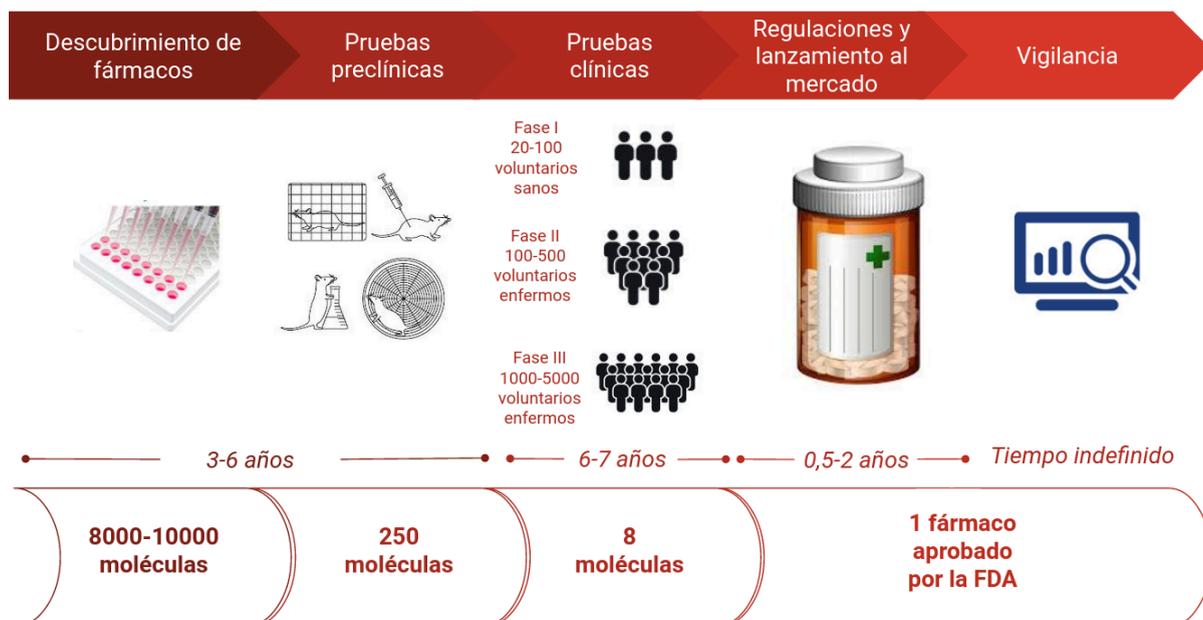


Figura 2.1. Esquema general del proceso de descubrimiento y desarrollo de un fármaco. Adaptado de Alberca, Tesis doctoral (2018).

La fase inicial del desarrollo de fármacos, el descubrimiento de fármacos, implica la identificación y validación del blanco molecular, así como la búsqueda de sustancias químicas que puedan modular dicho blanco. Las técnicas de Cribado de Alto Rendimiento (**HTS**, del inglés *High Throughput Screening*) permiten evaluar con rapidez decenas a cientos de miles de compuestos. Los compuestos que demuestran actividad satisfactoria frente al blanco molecular se denominan *hits*. A continuación, los *hits* se optimizan para mejorar su potencia y otras propiedades relevantes a través de ciclos de evaluación, lo que conduce a los "compuestos líderes" que pueden ser posteriormente optimizados. En esta etapa también se pueden utilizar estrategias *in silico* para determinar las características que deberían tener los análogos del *hit* o del líder para mejorar su actividad. Para guiar la optimización molecular, se realizan estudios en modelos animales (ensayos preclínicos) para evaluar la potencia y la toxicidad en sistemas complejos, y comenzar la caracterización farmacocinética *in vivo*. Finalmente, una vez que se han cumplido ciertos estándares de desempeño, el mejor candidato se somete a ensayos clínicos para determinar la seguridad, los efectos adversos, las dosis adecuadas, la eficacia y las propiedades farmacocinéticas y farmacológicas en humanos (Macalino et al., 2015; Hughes et al., 2011).

2.1.1 DESCUBRIMIENTO DE FÁRMACOS PARA ETDs

Las **ETDs**, como ya se ha mencionado en el capítulo anterior, son enfermedades que afectan principalmente a poblaciones en condiciones de vulnerabilidad socio-económica de países de ingresos bajos y medios, donde la atención sanitaria a estas afecciones suele ser deficitaria. Por éste último motivo, es importante contar con elementos que ayuden a la prevención y tratamiento de las **ETDs**, como por ejemplo fármacos o vacunas para la población afectada. Desafortunadamente, la cantidad de nuevos tratamientos para las **ETDs** es drásticamente inferior en comparación con otro tipo de afecciones, dando poco lugar a la innovación (Weng et al., 2018). Sin embargo, el descubrimiento de fármacos para **ETDs** ha incorporado estrategias que incluyen desde la química medicinal, pasando por la biología estructural, el modelado molecular, hasta ensayos tanto *in vitro* como *in vivo* utilizando organismos genéticamente modificados (Ferreira et al., 2022).

Las estrategias para el descubrimiento de fármacos para **ETDs** pueden clasificarse en las siguientes categorías:

1. **Estrategias basadas en el blanco molecular:** Consiste en cribar una biblioteca de compuestos generada contra un blanco molecular validado y la posterior optimización de los *hits* para aumentar la potencia y selectividad frente al blanco, como así también la farmacocinética del mismo (Moffat et al., 2017; Kiriiri et al., 2020).
2. **Estrategias basadas en el fenotipo:** Consiste en la evaluación de los compuestos en un sistema biológico, como puede ser una línea celular,

tejidos aislados o algún modelo animal. En este caso, no es necesario conocer a priori el o los mecanismos de inhibición del potencial *hit* obtenido (Moffat et al., 2017; Kiriiri et al., 2020).

3. **Estrategias basadas en el reposicionamiento de fármacos:** Se basa en la utilización de fármacos ya aprobados para otras patologías como potenciales agentes terapéuticos en la patología de interés (Kiriiri et al., 2020).

Estas tres estrategias se pueden complementar con métodos *in silico* para la reducción de los compuestos a evaluar, como así también para realizar una optimización racional de los *hits*.

Las estrategias basadas en el reposicionamiento de fármacos presentan como ventajas, por un lado, la reducción en los tiempos de aprobación del nuevo tratamiento por parte de organismos regulatorios, como así también una disminución en los costos asociados en la investigación y el desarrollo, en comparación con las otras dos estrategias mencionadas recientemente (Mignani et al., 2016). Se estima que el tiempo empleado en investigación y desarrollo para una nueva entidad química por las estrategias tradicionales oscila entre los 12 a 15 años (Macalino et al., 2015; Terry & Lesser, 2015), mientras que en el caso de una nueva indicación para un fármaco reposicionado dicho tiempo se reduce, llevándolo entre 3 a 12 años. Asimismo, los costos estimados para el desarrollo de una nueva entidad química se estiman entre 4-10 mil millones de dólares estadounidenses, en contraposición a los 1,6 mil millones de dólares estadounidenses para un fármaco reposicionado (Rudrapal et al., 2020; The Investopedia Team, 2022). Por otra parte, en base a las estrategias basadas en el blanco y en el fenotipo, se han realizado estudios que comparan el éxito de una y de otra en cuanto al descubrimiento de fármacos en la industria farmacéutica.

Swinney (2013) analizó los fármacos aprobados por la Administración de Alimentos y Medicamentos de los Estados Unidos (en inglés *Food and Drug Administration, FDA*) entre 1999 y 2008, discriminando entre aquellos obtenidos por estrategias basadas en el blanco o en estrategias basadas en el fenotipo. En dicho estudio, el autor describe el porcentaje de moléculas tipo fármaco *first-in-class* (primeros de su tipo, es decir inhibidores de una vía novedosa) y *follow-on* (moléculas que actúan por un mecanismo de acción ya conocido) aprobados para una y otra estrategia. Los datos muestran que las estrategias basadas en el fenotipo consiguieron una mayor cantidad de moléculas (28) tipo *first-in-class* que las estrategias basadas en el blanco (17), mientras que éstas últimas tuvieron un mejor rendimiento para fármacos tipo *follow-on* (83 compuestos versus 30 provenientes de ensayos fenotípicos). El autor hipotetiza que este bajo rendimiento en las estrategias basadas en el blanco se debe a la falta de consideración del entorno molecular de un sistema complejo, situación tenida en cuenta por los ensayos fenotípicos.

Un estudio posterior de Eder et al. (2014) extendió el período de análisis efectuado por Swinney (Swinney, 2013), abarcando a todas las entidades farmacológicas, incluyendo productos biofarmacéuticos, aprobadas en el período 1999-2013. En dicho período, de las 113 moléculas *first-in-class* aprobadas por la **FDA**, un 70% fueron descubiertas mediante estrategias basadas en el blanco, frente al 30% mediante estrategias basadas en el fenotipo. En el primer caso, 45 de los compuestos fueron tipo fármaco mientras que los 33 restantes son compuestos de origen biológico, es decir productos obtenidos a partir de técnicas de biología molecular, como por ejemplo anticuerpos monoclonales específicos para un blanco molecular determinado. Por otro lado, para las estrategias basadas en el fenotipo, la mayor parte de las moléculas *first-in-class* fueron identificadas mediante ensayos químico-céntricos (aquellos en donde el compuesto líder fue descubierto por similitud estructural a otro compuesto con actividad reportada contra uno o más blancos moleculares) y una pequeña parte por cribado fenotípico (es decir, el ensayo de un gran número de compuestos en el cual se evalúan cambios en el fenotipo de un organismo).

2.2 CRIBADO VIRTUAL

El cribado o tamizado virtual (**CV**), comprende un conjunto de técnicas computacionales cuyo objetivo es el cribado de grandes bases de datos con la finalidad de identificar potenciales compuestos químicos candidatos a fármaco (Walters et al., 2018). Talevi y Bruno-Blanch definen al **CV** como “*conjunto de técnicas computacionales que permiten, a partir de representaciones de la estructura molecular de los compuestos químicos almacenados en grandes bases de datos, identificar compuestos potencialmente interesantes desde el punto de vista farmacológico*” (Talevi & Bruno-Blanch, 2009).

En contraste con el **HTS**, el cual requiere de tecnología con altos costos de mantenimiento e infraestructura (Cronk, 2013), el **CV** ofrece una serie de ventajas que permiten su implementación en laboratorios de países en vías de desarrollo:

1. **Menor inversión necesaria:** En la actualidad, existen una gran cantidad de bases de datos públicas de compuestos químicos, como por ejemplo ChEMBL (Davies et al., 2015; Gaulton et al., 2017), la cual compila más de 2,3 millones de compuestos químicos y más de 20 millones de datos de actividad biológica (abril 2023); PubChem (Kim et al., 2016), con más de 115 millones de moléculas y más de 304 millones de datos de bioactividad (abril 2023), DrugBank (Law et al., 2014; Wishart et al., 2018), la cual compila alrededor de 9000 moléculas tipo fármaco, ZINC20 (Irwin et al., 2020; Sterling & Irwin, 2015; Irwin et al., 2012), que contiene más de 750 millones de estructuras químicas. A su vez, hay una amplia disponibilidad de softwares de uso libre o licencias académicas a bajo costo para las distintas etapas de

modelado en **CV**, tales como Mordred (Moriwaki et al., 2018), Chimera (Pettersen et al., 2004), OpenBabel (O'Boyle et al., 2011), RDKit (Rational Discovery LLC, 2006) ó MolVS (Swain, 2019). A su vez, existen servidores en línea de acceso libre que permiten realizar campañas de **CV** (González-Díaz et al., 2011; Schneidman-Duhovny et al., 2008; Liu et al., 2010; Konc et al., 2015). Por último, dependiendo de la estrategia de **CV** a utilizar y la cantidad de compuestos a cribar, es posible llevarlo adelante con computadoras de uso personal, con lo que se puede notar que la inversión en capital es infinitamente inferior en comparación al costoso **HTS**.

2. **Carácter teórico:** Dada la naturaleza *in-silico* del **CV**, no es necesario disponer físicamente de los compuestos a cribar. El **CV** permite el tamizado de compuestos químicos que aún no han sido sintetizados. En consecuencia, sólo serán adquiridas las moléculas con mejor desempeño en los modelos de **CV** (Talevi et al., 2011).
3. **Reposicionamiento de fármacos:** Existen varias bases de datos públicas que cuentan con fármacos aprobados, discontinuados o que se encuentran en estudios clínicos, como la ya mencionada DrugBank (Law et al., 2014; Wishart et al., 2018) y Sweetlead (Novick et al., 2013), con lo cual es posible tamizarlas por **CV** para encontrar nuevas indicaciones terapéuticas para fármacos ya existentes.
4. **Ventajas bioéticas:** La metodología de **CV** permite la reducción de ensayos *in vitro* e *in vivo* a realizar, cumpliendo con las pautas de bioética internacionales vigentes (Council for International Organization of Medical Sciences & The International Council for Laboratory Animal Science, 2012; Talevi & Bruno-Blanch, 2009).

En la **Figura 2.2** se puede observar el uso de metodologías aplicadas a **CV** en el período comprendido entre el 2004-2023 utilizando el buscador PubMed (National Institutes of Health, 2024). Se utilizó como palabra clave “*virtual screening*” en resúmenes y títulos de trabajos científicos entre 2004-2023.

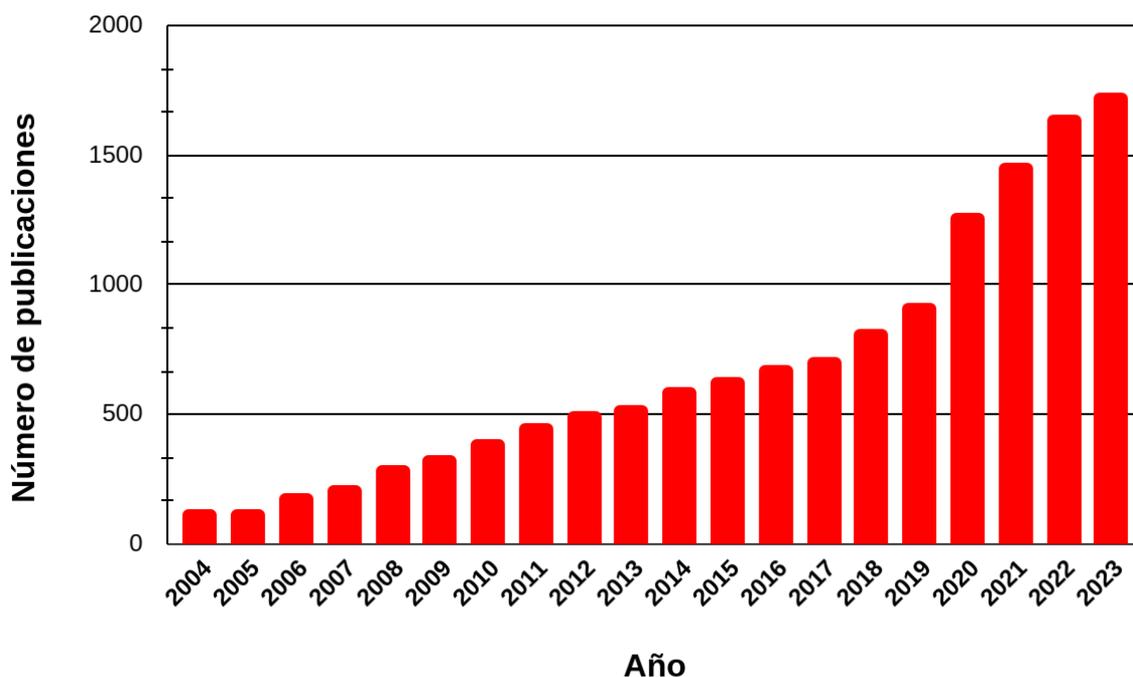


Figura 2.2. Número de artículos publicados en revistas científicas indexadas por PubMed que contienen el término “*virtual screening*” en su título y/o resumen en el período 2004-2023.

En términos generales, las estrategias para el diseño racional de fármacos asistido por computadora se clasifican en: Métodos basados en el ligando (o métodos indirectos) y Métodos basados en el blanco molecular o en la Estructura (o métodos directos). Ambas estrategias pueden ser combinadas, debido a su complementariedad, presentando cada una de ellas diferentes fortalezas y debilidades (Talevi et al., 2011; Sotriffer, 2009).

2.2.1 MÉTODOS BASADOS EN EL BLANCO MOLECULAR

Como su nombre lo indica, en los métodos basados en el blanco molecular es necesario conocer de antemano la estructura tridimensional de la biomolécula que se desea modular farmacológicamente. Para ello, la estructura tridimensional de las proteínas se pueden obtener por métodos experimentales (Ilari & Savino, 2008; Guerry & Herrmann, 2011) o computacionales (Talevi et al., 2011; Sotriffer, 2009).

Dentro de los métodos experimentales de determinación de estructura tridimensional de proteínas se pueden citar:

1. ***Cristalografía de Rayos X:*** Consiste en hacer incidir un haz de luz de longitud de onda correspondiente a los rayos X sobre un cristal de proteínas.

El patrón de difracción generado depende de las posiciones de los átomos en el cristal, posibilitando la determinación estructural de la molécula en cuestión (Gavernet, 2021).

2. **RMN**: Se basa en las propiedades de ciertos átomos que presentan *spin* nuclear diferente de cero (^1H , ^{13}C , ^{15}N). Al aplicar un campo magnético sobre la muestra, dichos núcleos se alinearán a favor (mayoritariamente) o en contra de dicho campo. Esta diferencia energética entre ambos estados permite la aplicación de pulsos de radiación externa para conseguir la transición hacia el estado de mayor energía. Cuando la energía del pulso es cuantitativamente igual a dicha barrera energética, se produce una absorción, y consecuentemente, una señal medible. Debido a que la frecuencia de absorción de cada núcleo depende del tipo de átomo y de su entorno químico, es posible reconstruir la estructura tridimensional de una molécula (Gavernet, 2021).
3. **Crio-microscopía electrónica (en inglés Cryo-EM)**: En esta técnica, se utiliza a la microscopía electrónica para la determinación de estructuras moleculares. Brevemente, este tipo de microscopía utiliza haces de electrones, que al impactar con la muestra, generan la dispersión en la trayectoria de los mismos. Esta dispersión es dependiente de la estructura del material analizado. Como paso previo, debe congelarse la muestra a temperaturas criogénicas, lo cual permite, por un lado, la conservación de la estructura molecular a determinar y, por otro, la minimización del daño al material de estudio por efectos de la radiación del equipo (Carroni & Saibil, 2016). Finalmente, la microscopía electrónica genera imágenes 2D correspondientes a diferentes proyecciones de la molécula en la muestra. Una vez generadas dichas proyecciones, es posible obtener la reconstrucción de un modelo 3D a partir de programas de procesamiento de imágenes (Nogales & Scheres, 2015).

Por su parte, los métodos computacionales de determinación de estructura terciaria de proteínas se pueden clasificar en dependientes de la estructura molde (del inglés *template-based modeling*) e independientes de la estructura molde (del inglés *template-free modeling*).

1. **Métodos dependientes de la estructura molde**: En este tipo de métodos, es necesaria la comparativa entre la secuencia aminoacídica del blanco de interés y al menos una secuencia de otra proteína con estructura tridimensional conocida, la cual se conoce como molde (Fiser, 2010). Dentro de esta categoría podemos describir dos metodologías:
 - **Modelado por homología**: Se basa en el hecho de que dos ó más secuencias de aminoácidos cuyo porcentaje de identidad es superior al 30%, tienden a exhibir una estructura tridimensional similar (Rost, 1999). Consecuentemente, al conocer la estructura de al menos una de estas secuencias, esta información puede ser transferida a aquellas

proteínas cuyo arreglo tridimensional no ha sido determinado aún (Gavernet, 2021).

- **Modelado por threading o reconocimiento de plegamiento:** Consiste en la asignación de las coordenadas de una estructura determinada experimentalmente a la secuencia problema y la posterior evaluación de dicha asignación mediante la optimización de una función de puntuación (Peng & Xu, 2009). A diferencia del modelado por homología, este método permite identificar proteínas con un mismo plegamiento, sin que existiera un origen evolutivo común entre ellas (Gavernet, 2021).

2. **Métodos independientes de la estructura molde:** Este tipo de métodos permiten obtener la estructura tridimensional de una proteína contando únicamente con su estructura primaria. Dentro de este inciso, se pueden encontrar dos metodologías:

- **Modelado ab-initio:** Generan un modelo estructural con la información obtenida a partir de la secuencia de aminoácidos de la proteína cuya estructura se quiere determinar. Los programas utilizan algoritmos que exploran las posibles conformaciones que puede adoptar la secuencia mediante la búsqueda de mínimos energéticos. Aunque este tipo de modelados no presenta una limitación asociada a los plegamientos ya conocidos, las funciones de energía que se utilizan para describir el plegado proteico no son lo suficientemente exactas para generar resultados precisos (Gavernet, 2021).
- **Modelado por aprendizaje automático:** Se basa en la idea de que la estructura de una proteína está determinada por su secuencia de aminoácidos, y que los patrones en las secuencias de proteínas pueden ser utilizados para predecir su estructura tridimensional. El aprendizaje automático es una rama de la inteligencia artificial que se enfoca en el desarrollo de algoritmos y modelos que pueden aprender patrones en grandes conjuntos de datos y utilizar esta información para hacer predicciones en nuevos datos. En el contexto del modelado estructural de proteínas, se utilizan técnicas de aprendizaje automático, como redes neuronales, para aprender a predecir la estructura de proteínas a partir de la información disponible en grandes bases de datos de proteínas, como puede ser por ejemplo el **PDB** (Jumper et al., 2021). En la siguiente sección se describe con mayor detalle esta última metodología, por ser la utilizada en esta tesis doctoral.

Los métodos desde el blanco molecular pueden ser utilizados con dos puntos de vista diferentes, por un lado en el diseño de fármacos para favorecer interacciones entre el blanco molecular y compuestos de interés, o bien en el **CV** de bases de datos, en donde se busca estimar la energía de interacción entre los compuestos

presentes en la base de datos y la diana farmacológica de interés. Estos dos enfoques complementarios serán abordados en mayor detalle en la sección 2.2.1.2.

2.2.1.1 MODELADO POR APRENDIZAJE AUTOMÁTICO

Desde la publicación del algoritmo de AlphaFold2 en el año 2020, el paradigma del modelado estructural de proteínas ha cambiado por completo. AlphaFold2 es un sistema de inteligencia artificial basado en el aprendizaje profundo, cuyo objetivo es predecir la estructura tridimensional de una proteína a partir de su secuencia de aminoácidos, siendo muy útil en el campo de la investigación biomédica (Porta-Prado et al., 2022).

La predicción de la estructura de las proteínas es un problema muy difícil de resolver, debido a la complejidad de las interacciones entre los aminoácidos y los diferentes factores que influyen en la estructura de las proteínas (The AlphaFold team, 2020). Los métodos tradicionales para predecir la estructura de las proteínas, como se mencionó en la sección anterior, necesitan de estructuras relacionadas a la secuencia problema que estén presentes en el **PDB** para llevar adelante el modelado comparativo (Fiser, 2010), o bien dependen de la minimización de un potencial energético, el cual no siempre es lo suficientemente preciso para describir adecuadamente la estructura (Gavernet, 2021).

El desempeño de AlphaFold ha sido evaluado en varias competiciones de predicción de estructuras de proteínas, conocidas como competiciones **CASP** (acrónimo de *Critical Assessment of Techniques for Protein Structure Prediction*). En ellas, AlphaFold ha superado de manera consistente a otros sistemas de predicción de proteínas. La primera versión de AlphaFold, denominada AlphaFold1, obtuvo el primer lugar en la clasificación general de la 13^a edición de la competición **CASP** en diciembre de 2018. Esta versión se destacó en la predicción de estructuras para las cuales no existían modelos previos, lo cual fue considerado por los organizadores de la **CASP** como uno de los mayores desafíos (The AlphaFold team, 2022).

En 2020, se desarrolló la segunda versión de AlphaFold, denominada AlphaFold2, la cual volvió a ganar la competición **CASP** en noviembre de ese año. Las predicciones realizadas por AlphaFold2 fueron mucho más precisas que las de cualquier otro programa. En aproximadamente dos tercios de las proteínas evaluadas, AlphaFold2 obtuvo una puntuación superior a 90 en la prueba de distancia global (**GDT**, acrónimo de *global distance test*), que compara las predicciones de los programas con las estructuras determinadas experimentalmente (Douglas Heaven, 2020; Service, 2020). Cabe destacar que una puntuación de 100 indica una coincidencia perfecta entre la predicción y la estructura experimental. Aunque en un tercio de los casos la precisión de las predicciones no fue lo suficientemente alta y el programa no proporcionó información sobre el mecanismo

del plegamiento de las proteínas (Curry, 2020; Balls, 2020), el logro técnico de AlphaFold ha sido ampliamente reconocido.

Brevemente, la predicción de estructuras basada en aprendizaje profundo se enfoca en reconocer mapas de contacto entre residuos y matrices de distancia entre residuos, con el objetivo de encontrar patrones en las secuencias y estructuras de las proteínas en **PDB** (Kuhlman & Bradley, 2019).

Basándose en este principio, AlphaFold2 es capaz de predecir estructuras tridimensionales para todos los átomos pesados de una secuencia dada utilizando un alineamiento múltiple de secuencias (**MSA**) para predecir las distancias entre los átomos dentro de la secuencia a modelar (Jumper et al., 2021). Utilizando una secuencia proporcionada por el usuario, AlphaFold2 implementa los métodos HMMer y HHblits para recuperar secuencias homólogas de UniRef100 y otras bases de datos. Además, AlphaFold2 también intenta encontrar estructuras molde para construir una representación inicial de la proteína modelada, conocida como representación de pares, que proporciona una hipótesis inicial sobre qué aminoácidos es probable que estén en contacto entre sí. Posteriormente, AlphaFold2 direcciona al **MSA** y la representación de pares hacia una serie de capas de red profunda conocidas como capas *Evoformer*. Estas capas extraen las características más relevantes del **MSA** y la representación de pares de manera iterativa e intercambian información entre ellas. Finalmente, la salida de *Evoformer* se utiliza como entrada para el módulo de estructura, que produce un modelo 3D. Este procedimiento se repite tres veces de manera iterativa para mejorar la precisión del modelo predicho (Jumper et al., 2021).

Dado que AlphaFold2 requiere una gran cantidad de recursos computacionales, las secuencias de proteínas modeladas en esta tesis doctoral se enviaron al algoritmo de AlphaFold2 a través de la plataforma Google Colab utilizando los parámetros por defecto (The AlphaFold Team & Google Colab, 2022). Esta versión se llama ColabFold y reemplaza los métodos de búsqueda HMMer y HHblits por MMseq2, con el fin de reducir el tiempo de cómputo, así como los requisitos de memoria **RAM** (del inglés *random access memory*) y almacenamiento (Mirdita et al., 2021).

La **Figura 2.2.1.1** muestra un esquema simplificado acerca del funcionamiento del algoritmo de AlphaFold2.

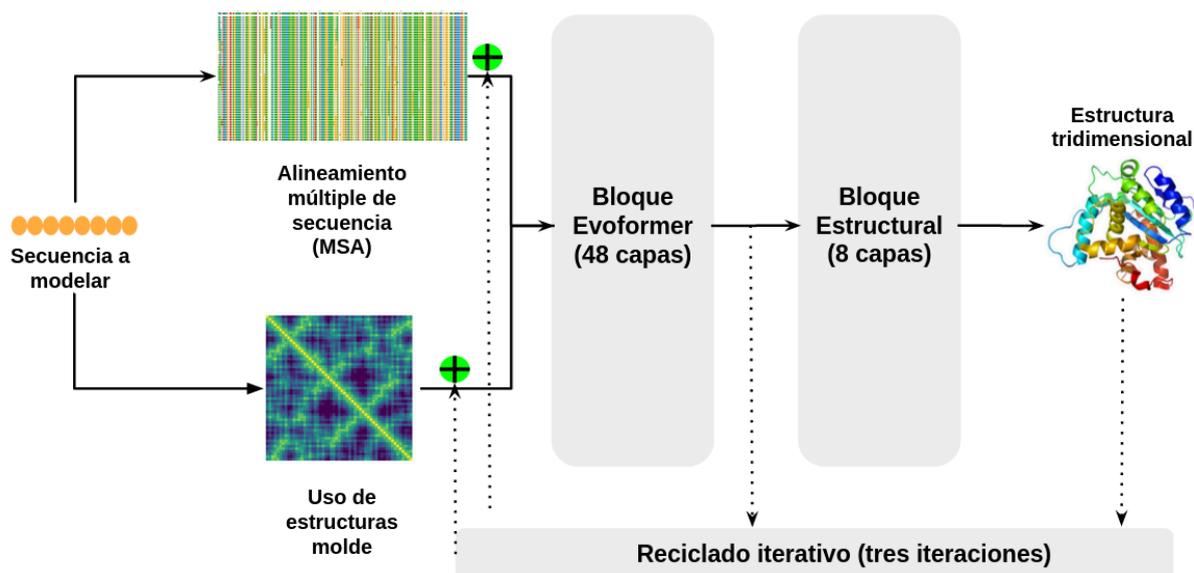


Figura 2.2.1.1. Esquema simplificado del algoritmo implementado en AlphaFold2. Adaptado de Jumper et al. (2021).

2.2.1.2 DOCKING MOLECULAR

El *docking* molecular es una técnica de simulación que tiene como objetivo predecir la orientación espacial relativa de un ligando en un receptor, así como asignar a cada orientación un puntaje que, en principio, reflejaría la afinidad o energía de unión asociada a la formación del complejo ligando-receptor, a partir de las estructuras tridimensionales de ambos componentes.

Los programas de *docking* se componen de dos algoritmos principales: uno que explora las distintas conformaciones del ligando (algoritmo de búsqueda) y otro que evalúa cada una de esas poses y les asigna un valor de *score* o energía de unión, basándose en las interacciones con el receptor y el conformero del ligando acoplado al sitio de unión definido (función de *score* o puntuación). Como resultado, el programa predice una serie de poses del ligando en el sitio de unión del receptor, ordenadas o rankeadas según su energía de unión predicha. Las poses mejor rankeadas son consideradas las más probables (Gaitonde et al., 2020). El ligando utilizado en el *docking* es generalmente una molécula orgánica pequeña, mientras que el receptor puede ser una proteína o un ácido nucleico (Luo et al., 2019). También es posible utilizar la técnica de *docking* en un sistema proteína-proteína (Porter et al., 2019).

El desempeño del *docking* puede evaluarse, por un lado, por su capacidad de reproducir correctamente el modo de unión de un ligando en un receptor (lo que se conoce en inglés como *docking power*) y, por otro, por la capacidad de la función de

puntuación de predecir correctamente la afinidad o energía de formación del complejo ligando-receptor (*scoring power* en inglés) (Gavernet, 2021).

El *docking power* puede ser evaluado mediante experimentos de *re-docking* y *cross-docking*. El *re-docking* implica extraer el ligando de un complejo ligando-receptor previamente determinado experimentalmente y luego volver a acoplarlo en el mismo sitio. El *cross-docking*, por otro lado, implica extraer el ligando del complejo y luego acoplarlo en otra estructura del receptor, ya sea en una estructura holo diferente o en una estructura apo. Para evaluar el rendimiento de la predicción, se compara la pose predicha con el modo de unión determinado experimentalmente utilizando alguna métrica, como la desviación cuadrática media (**RMSD**, del inglés *root mean square deviation*) (Kirchmair et al., 2008).

Por otro lado, el *scoring power* se refiere a la capacidad de una función de puntuación para predecir correctamente la afinidad o energía de formación del complejo ligando-receptor. Esto se puede evaluar mediante el cribado retrospectivo de un conjunto de validación. Dependiendo de la naturaleza de la actividad biológica de los compuestos en dicho conjunto, se pueden utilizar diferentes métricas. Si la variable dependiente es continua (como la concentración inhibitoria media, **IC₅₀** o la constante de inhibición, **Ki**), es posible calcular métricas de correlación o de ranking entre la afinidad predicha y la determinada experimentalmente (Llanos et al., 2020). Si la variable dependiente es binaria (por ejemplo, activo o inactivo), se puede evaluar el rendimiento del *docking* como un clasificador binario utilizando el valor de afinidad predicho. Los procedimientos tanto para la realización de un experimento de *docking power* como de *scoring power* son ilustrados esquemáticamente en la **Figura 2.2.1.2**.

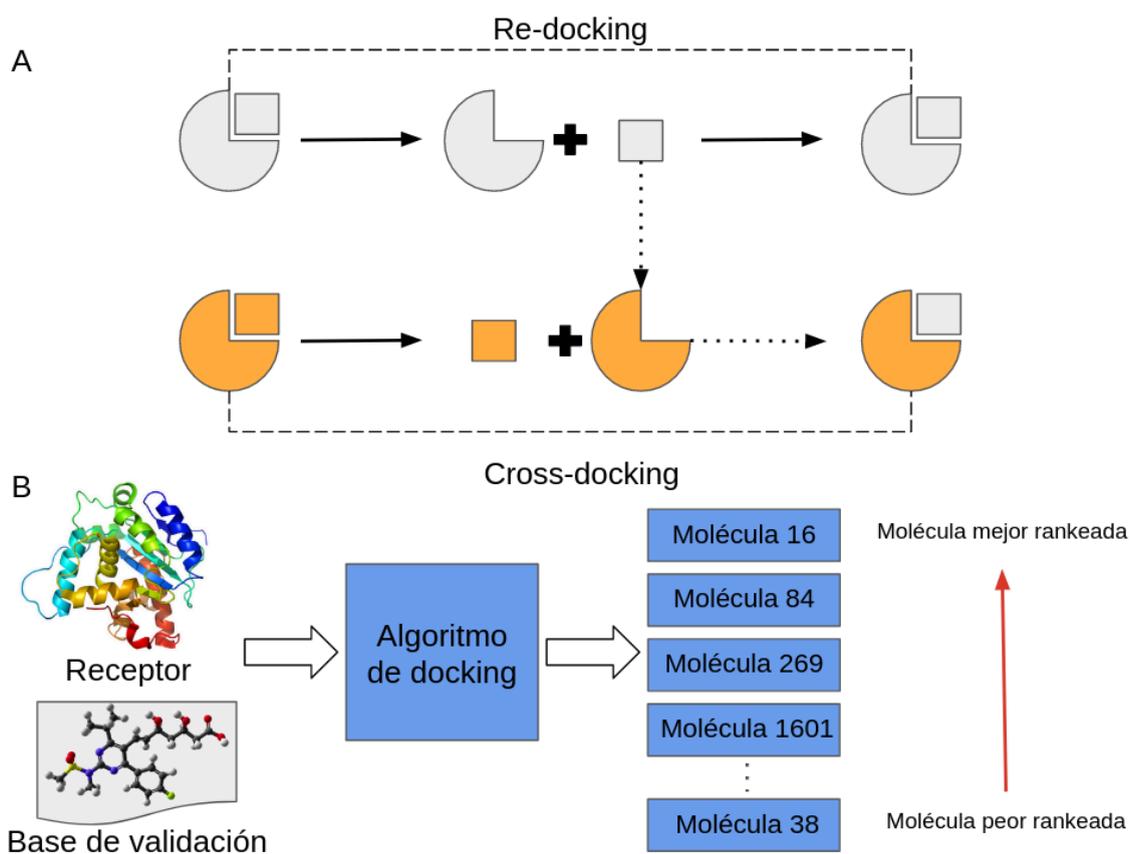


Figura 2.2.1.2. Esquema genérico de un experimento de *docking power* (A) y de *scoring power* (B).

En cuanto a la flexibilidad considerada para el ligando y el receptor durante la simulación, existen varias aproximaciones. En el *docking* rígido, ambos se consideran como cuerpos rígidos y solo se permite al ligando rotar y trasladarse dentro del sitio de unión. En el *docking* flexible, se exploran diferentes conformaciones del ligando, de la proteína o de ambos simultáneamente (Andrusier et al., 2008).

Existen diversas estrategias que permiten considerar la flexibilidad del blanco en las simulaciones de *docking* (Lexa & Carlson, 2012), como el uso de potenciales suavizados o *soft docking* (Ferrari et al., 2004), el uso de bibliotecas de rotámeros para modelar la flexibilidad de las cadenas laterales de ciertos residuos (Alberts et al., 2005), o considerar distintas conformaciones del receptor (*ensemble docking*) (Amaro et al., 2018). Sin embargo, a medida que se incluyen más grados de libertad en la simulación, el costo computacional aumenta significativamente. Por lo tanto, la aproximación más utilizada es considerar el ligando totalmente flexible y la proteína rígida (Meng et al., 2011).

2.2.1.3 DINÁMICA MOLECULAR

La técnica de simulación conocida como dinámica molecular (**DM**) permite el estudio del comportamiento dinámico de moléculas biológicas, orgánicas o inorgánicas a lo largo del tiempo (Karplus & McCammon, 2002). En la versión clásica de la **DM**, las diferentes configuraciones de un sistema (es decir, las posiciones relativas de todos los átomos que lo componen) se generan resolviendo las ecuaciones de movimiento. La **Ecuación 2.1** describe la segunda ley de Newton, en donde $\vec{F}(t)$ se corresponde con la fuerza ejercida sobre una partícula a un tiempo t dado, siendo proporcional al cambio del momento \vec{p} de dicha partícula (es decir, el producto entre su masa y velocidad) con respecto a t . La integración numérica de la **Ecuación 2.1** permite calcular la trayectoria de cada partícula, es decir, cómo cambia su posición y velocidad con el tiempo. Para esto, es necesario conocer las fuerzas que actúan sobre cada átomo del sistema. Dichas fuerzas derivan de la función de energía potencial (**Ecuación 2.2**) (Tuckerman & Martyna, 2000).

$$\vec{F}(t) = \frac{d\vec{p}}{dt} \quad (2.1)$$

$$E(r^N) = E_{enlaces} + E_{ángulos} + E_{diedros} + E_{no\ enlazantes} \quad (2.2)$$

$$E_{enlaces} = \sum_{enlaces} K_r (r - r_0)^2 \quad (I)$$

$$E_{ángulos} = \sum_{ángulos} K_\theta (\theta - \theta_0)^2 \quad (II)$$

$$E_{diedros} = \sum_{diedros} \left(\frac{V_1}{2} [1 + \cos(\phi - \phi_1)] + \frac{V_2}{2} \right)$$

$$[1 - \cos(2\phi - \phi_2)] + \frac{V_3}{2} [1 + \cos(3\phi - \phi_3)] + \frac{V_4}{2} [1 - \cos(4\phi - \phi_4)] \quad (III)$$

$$E_{no\ enlazantes} = \sum_{i>j} f_{ij} \left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^6} + \frac{q_i q_j e^2}{4\pi\epsilon_0 r_{ij}} \right) \quad (IV)$$

En la **Ecuación 2.2**, se puede ver que la energía potencial del sistema, $E(r^N)$, depende de una serie de parámetros específicos relacionados con cada uno de los términos que conforman el campo de fuerza (expresiones **I-IV**). Estos parámetros se obtienen a partir de datos experimentales y cálculos teóricos, y su ajuste es crucial para lograr una descripción precisa del sistema molecular. Las contribuciones de las interacciones enlazantes incluyen a la energía de los enlaces ($E_{enlaces}$), la energía asociada a las oscilaciones angulares que comprenden a tres átomos consecutivos ($E_{ángulos}$) y la energía de los ángulos diedros ($E_{diedros}$). Por su parte, las interacciones no enlazantes ($E_{no\ enlazantes}$) incluyen a las interacciones del tipo van der Waals y coulombicas, representadas por la expresión **IV** de la **Ecuación 2.2**. En este trabajo de tesis, el campo de fuerza utilizado fue OPLS-AA, implementado dentro del programa de **DM** GROMACS (Jorgensen et al., 1996; Berendsen et al., 1995; Lindahl et al., 2001; Van Der Spoel et al., 2005; Hess et al., 2008; Pronk et al., 2013; Páll et al., 2015; Abraham, et al., 2015).

La simulación de **DM** se puede llevar a cabo en distintas condiciones de simulación. El ensamble microcanónico (**NVE**) es útil para estudiar sistemas aislados, debido a que el número de partículas (**N**), el volumen (**V**) y la energía (**E**) del sistema permanecen constantes (Yao et al., 2022). Por otro lado, el ensamble canónico (**NVT**) permite fluctuaciones de energía dentro de los límites del sistema a través del intercambio de calor. Por lo tanto, es menester la presencia de un algoritmo capaz de controlar las fluctuaciones de temperatura dentro de ciertos límites, simulando un baño termostatizado. Finalmente, el ensamble isotérmico-isobárico (**NPT**) permite simular sistemas a presión y temperatura constantes, situación habitual en un laboratorio de química. Para ello, además de controlar la temperatura mediante un termostato, es necesario también que un algoritmo controle la presión, es decir, un barostato (Yao et al., 2022).

Un protocolo de **DM** consta típicamente de tres etapas principales: preprocesamiento, simulación y análisis de resultados. Durante la etapa de preprocesamiento, se prepara el sistema molecular para la simulación, se elige el campo de fuerza adecuado y se establecen las condiciones de simulación. Durante la etapa de simulación, primeramente, se equilibra el sistema de acuerdo al tipo de ensamble elegido. Una vez alcanzada la convergencia en las variables del sistema sometidas a equilibración (por ejemplo, la temperatura absoluta), se procede a la producción de **DM**, en donde se lleva a cabo un muestreo de las distintas configuraciones del sistema. Finalmente, durante la etapa de análisis de resultados, se evalúan las propiedades y características del sistema simulado (Katiyar & Jha, 2018).

La simulación de **DM** es una herramienta muy útil para el estudio de sistemas moleculares, permitiendo considerar al blanco molecular de interés en un entorno similar al del experimento *in vitro*, es decir rodeado de solvente a la temperatura, presión y fuerza iónica de interés, lo cual permite un modelado más preciso de las propiedades fisicoquímicas del sistema (Hollingsworth & Dror, 2018).

2.2.2 MÉTODOS BASADOS EN EL LIGANDO

En contraposición con los métodos directos, el enfoque implementado en esta sección tiene como principal hipótesis que moléculas similares tendrán propiedades de unión similares a un blanco molecular dado (Sotriffer, 2009).

Los métodos basados en el ligando utilizan la información proveniente de un ligando activo o un grupo de ligandos activos contra el blanco molecular de interés para identificar características estructurales responsables de la actividad biológica observada (Macalino et al., 2015). Dependiendo de la técnica de **CV** que se desee utilizar, los ligandos conocidos pueden caracterizarse mediante la representación bidimensional (2D) y/o tridimensional (3D) de sus estructuras químicas, con el objetivo de encontrar determinantes estructurales de la actividad biológica.

En los siguientes apartados de este capítulo, se describirán brevemente los métodos de **CV** basados en el ligando, que se pueden clasificar en tres grandes grupos: enfoques basados en el farmacóforo, metodologías basadas en la similitud molecular y metodologías basadas en descriptores moleculares.

2.2.2.1 FARMACÓFOROS

Según la **IUPAC**, un farmacóforo se trata de un grupo de propiedades estéricas y electrónicas que aseguran las interacciones supramoleculares óptimas con un blanco biológico específico, activando o bloqueando su respuesta (Wermuth et al., 1998). Por otro lado, Talevi et al. (2011) definen a un farmacóforo como “una

configuración tridimensional que posee un conjunto mínimo de características moleculares necesarias para garantizar interacciones con un objetivo molecular específico, generando una respuesta biológica determinada”.

El **CV** basado en farmacóforos implica la superposición de un conjunto de moléculas que comparten una actividad específica (conjunto de entrenamiento), con el fin de extraer las propiedades químicas comunes que son esenciales para producir la actividad de interés. Posteriormente, se buscan en bases de datos los compuestos que presenten el farmacóforo obtenido (Sotriffer, 2009).

Por lo general, la obtención de la estructura farmacofórica se realiza mediante la extracción de las características comunes de las estructuras 3D del conjunto de entrenamiento. Sin embargo, para construir y aplicar hipótesis farmacofóricas, es menester tener en cuenta, por un lado, la flexibilidad de las moléculas, y por otro, cómo llevar adelante la superposición molecular (Yang, 2010).

Con respecto a la flexibilidad de las moléculas, tradicionalmente se han utilizado diversas estrategias, como por ejemplo la utilización de uno o más ligandos rígidos, la generación de múltiples conformaciones posibles de los compuestos del conjunto de entrenamiento y de la base sometida al cribado; y el análisis conformacional continuo durante el proceso de modelado farmacofórico (Talevi et al., 2011). Sin embargo, en los últimos años, es cada vez más frecuente la integración de **DM** en protocolos de **CV** basado en farmacóforos, dado que permite aliviar los problemas asociados a la flexibilidad del ligando (Muhammed & Aki-Yalcin, 2021). En cuanto a la superposición molecular, el alineamiento se puede llevar a cabo utilizando enfoques basados en puntos, en los cuales se minimizan las distancias entre átomos o puntos de características químicas abstractas, o mediante aproximaciones basadas en propiedades, que utilizan descriptores basados en campos moleculares para generar el alineamiento (Talevi et al., 2011). La primera opción es la más utilizada, ya que un arreglo tridimensional de características moleculares abstractas en lugar de grupos funcionales bien definidos permite recuperar una mayor diversidad estructural en los compuestos activos. Estas características suelen incluir grupos con capacidad de establecer enlaces de hidrógeno, grupos cargados, grupos capaces de participar en contactos hidrofóbicos o aromáticos, y pueden considerarse restricciones estéricas bajo la forma de volúmenes de exclusión (Kaserer et al., 2015). También se puede abordar la identificación y utilización de farmacóforos desde el receptor, buscando posibles puntos de interacción entre el blanco molecular y los ligandos (Yang, 2010).

2.2.2.2 SIMILITUD MOLECULAR

En el caso de disponer de una escasa cantidad de ligandos con actividad conocida contra un blanco molecular, se pueden utilizar técnicas basadas en la similitud molecular como estrategia de **CV** (Talevi et al., 2011). Estas técnicas implican la

medición del grado de similitud entre la estructura de un compuesto activo deseado y los compuestos de una base de datos sometida a cribado (Bajorath, 2001). La similitud molecular se enfoca principalmente en las características estructurales de los compuestos, como la presencia, ausencia o frecuencia de determinadas subestructuras químicas o grupos funcionales. Maggiora et al. (2014) postulan que, a la hora de cuantificar la similitud entre entidades moleculares, se requieren tres elementos clave:

1. **Una representación molecular adecuada:** Es decir, una manera de representar las características moleculares de relevancia, como puede ser la utilización de descriptores moleculares.
2. **Un sistema de ponderación para las características moleculares:** En otras palabras, consiste en definir si la contribución de las características moleculares definidas es la misma para estimación de la similitud o por el contrario, si hay características con mayor peso que otras.
3. **Un coeficiente de similitud:** Esto es, una expresión matemática que transforme en un valor numérico la información contenida en la representación molecular. Usualmente, este valor se encuentra entre 0 y 1, donde 1 se corresponde con la identidad completa de la representación molecular. Sin embargo, esto no implica que los compuestos comparados sean idénticos.

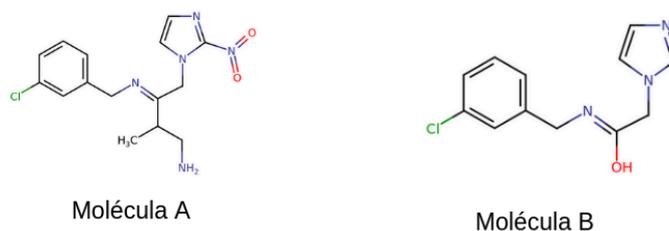
Una forma común de comparar cuantitativamente dos moléculas se basa en la comparación de secuencias de *bits*, también conocidas como huellas digitales moleculares o *fingerprints* en inglés. Las huellas digitales moleculares se construyen asignando cada *bit* a la presencia o ausencia de una propiedad estructural determinada, que puede ser bidimensional o tridimensional. Alternativamente, los *bits* pueden indicar la frecuencia con la que aparece la propiedad estructural en las moléculas comparadas. Los *bits* compartidos por ambas moléculas se cuantifican mediante el coeficiente de similitud, y uno de los coeficientes de similitud más utilizados es el coeficiente de Tanimoto en su versión binaria (Willett, 2006), que se define en la **Ecuación 2.3**:

$$S = \frac{c}{a + b - c} \quad (2.3)$$

Las letras a y b simbolizan el número de *bits* correspondientes a las moléculas A y B a analizar, respectivamente, mientras que c representa el número de *bits* comunes a ambas moléculas.

En la **Figura 2.2.2.2** se ilustra un ejemplo de cómo se lleva a cabo el cálculo del coeficiente de Tanimoto. Considerando las 8 subestructuras de la figura, se puede determinar que el número de *bits* para la molécula A es de 7, mientras que el número de *bits* para la molécula B es de 5. El número de subestructuras

compartidas entre ambas moléculas es de 4. En consecuencia, el coeficiente de Tanimoto se establece en 0,5.



Molécula A	1	1	1	1	0	1	1	1
Molécula B	1	0	1	0	1	1	0	1

Figura 2.2.2.2. Esquema del proceso de comparación de dos moléculas por medio de un coeficiente de similitud. Adaptado de Gavernet (2021).

Las técnicas de **CV** basadas en la similitud presentan dos ventajas principales. En primer lugar, tienen un bajo costo computacional, lo que les permite tamizar grandes bases de datos en un corto período de tiempo. En segundo lugar, permiten realizar búsquedas virtuales a partir de una única molécula de referencia (Di Ianni et al., 2015).

2.2.2.3 DESCRIPTORES MOLECULARES

Según Todeschini y Consonni (2009), un descriptor molecular “es el resultado de un proceso lógico y matemático que convierte información química en una representación simbólica de una molécula en un número útil o resultado de un experimento estandarizado”. Un descriptor molecular es una variable numérica que representa alguna característica de la estructura molecular, como el tamaño, la forma, la lipofilidad, etc. (Davis, 2014).

Los descriptores moleculares pueden variar en complejidad, desde propiedades simples como el peso molecular hasta descriptores geométricos de mayor complejidad de cálculo. La dimensionalidad de los descriptores se puede clasificar en 0D, 1D, 2D, 3D y 4D, según la información contenida en ellos.

Los descriptores 0D se derivan de la fórmula química de la molécula y son muy fáciles de calcular e interpretar, pero tienen bajo contenido de información y pueden tener valores iguales para diferentes moléculas. Por su parte, los descriptores 1D examinan la frecuencia u ocurrencia de determinadas subestructuras o grupos funcionales en la molécula, siendo los valores calculados de naturaleza binaria o

discreta (Grisoni et al., 2018). En el caso de los descriptores 2D, estos consideran la conectividad de los átomos y reflejan a la molécula como un grafo, derivando descriptores de la topología molecular. A su vez, también pueden reflejar la distribución en la molécula de propiedades atómicas (Pogliani, 2000; Consonni & Todeschini, 2012). En cuanto a los descriptores 3D, la molécula es representada como un objeto geométrico caracterizado por la configuración espacial de sus átomos y tienen un alto contenido de información, pero son computacionalmente más costosos que los descriptores 0-2D (Grisoni et al., 2018). Por último, los descriptores 4D son similares a los 3D, pero consideran la información estructural para un conjunto de conformeros.

A medida que aumenta la dimensionalidad de representación, se obtiene más información de la molécula, pero también aumenta el costo computacional asociado a los cálculos (Andrade et al., 2010). La representación gráfica de la molécula de clomipramina se muestra en la **Figura 2.2.2.3** mediante diferentes representaciones que varían en su dimensión. Se puede notar que a medida que aumenta la dimensionalidad de la representación, se obtiene más información sobre la molécula, aunque esto conlleva un aumento en el costo computacional asociado a los cálculos.

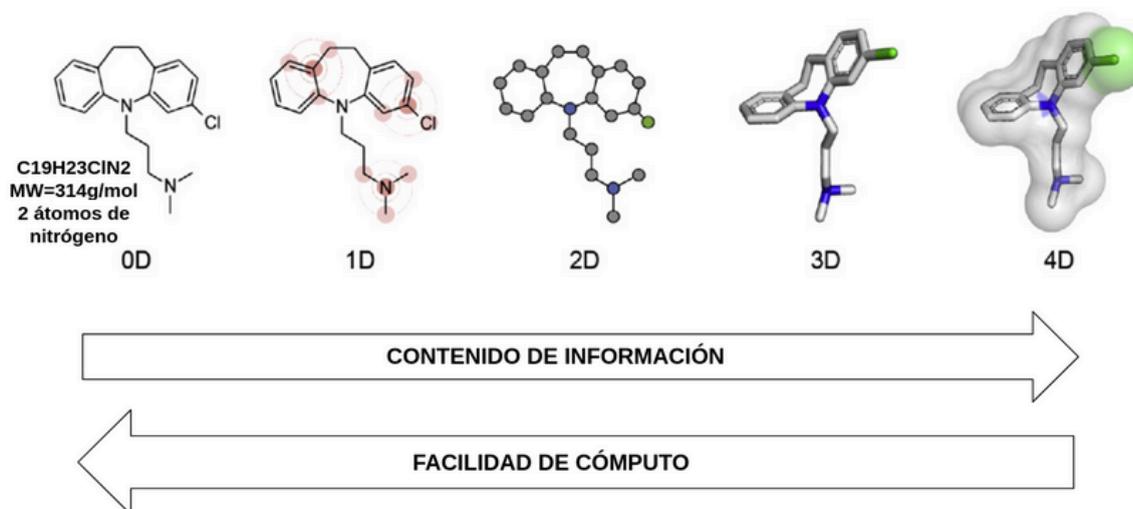


Figura 2.2.2.3. Distintas representaciones de la molécula de clomipramina. Adaptado de Fernández-Torras et al. (2022).

Desde un punto de vista práctico, los descriptores moleculares pueden clasificarse en dos categorías, de acuerdo a su dimensionalidad: los de baja dimensionalidad (0D, 1D, 2D) que son independientes de la conformación, y los de alta dimensionalidad (3D, 4D) que son dependientes de la conformación.

Aplicación de los descriptores moleculares al CV

Para encontrar correlaciones entre la actividad a modelar y los descriptores moleculares, se utilizan distintas estrategias para constituir las llamadas Relaciones Cuantitativas Estructura-Actividad (**QSAR**, por sus siglas en inglés) (Talevi & Bruno-Blanch, 2009; De et al., 2022). Estos modelos se basan en la noción de que la actividad biológica de un compuesto químico depende de sus características estructurales, siendo que variaciones estructurales en una serie de compuestos conducirán a variaciones en su(s) propiedad(es) biológica(s) (Macalino et al., 2015). Existen dos tipos de modelos **QSAR**: cuantitativo y cualitativo, dependiendo de la naturaleza de la variable dependiente. Los modelos **QSAR** cuantitativos predicen el valor numérico de la propiedad (**K_i**, **pIC₅₀**, **pED₅₀**, etc.), mientras que los modelos cualitativos generan una respuesta discreta asociada a diferentes categorías (activo o inactivo, tóxico o no tóxico, etc.). Una vez obtenido un modelo **QSAR**, puede aplicarse a campañas de **CV** para predecir la actividad de las estructuras químicas contenidas en quimiotecas digitales (Neves et al., 2018).

2.2.3 MÉTODOS COMBINADOS

Los métodos mixtos, híbridos o combinados son una estrategia para el tamizado molecular que utiliza una combinación de metodologías basadas en el ligando y en la estructura de forma secuencial o paralela (**Figura 2.2.3**). El **CV** jerárquico (**CVJ**) utiliza diferentes métodos *in silico* en orden creciente con respecto al costo computacional asociado de cada uno de ellos. De esta manera, las quimiotecas virtuales serán cribadas en primer lugar por un método de bajo coste computacional, y los compuestos candidatos de esta primera etapa serán sometidos a un segundo método que requiera un mayor poder de cómputo, y así sucesivamente, hasta alcanzar un número manejable de compuestos para los ensayos biológicos (Kumar & Zhang, 2015; Miles & Ross, 2021). Por otro lado, el **CV** paralelo (**CVP**) ejecuta simultáneamente varios métodos complementarios, y los mejores *hits* de cada método son seleccionados para las pruebas biológicas (Kumar & Zhang, 2015). La ventaja de la complementariedad en los métodos utilizados es que permite obtener, a priori, una mayor diversidad estructural en los compuestos candidatos obtenidos, aunque a expensas de un mayor costo computacional con respecto a **CVJ** (Miles & Ross, 2021).

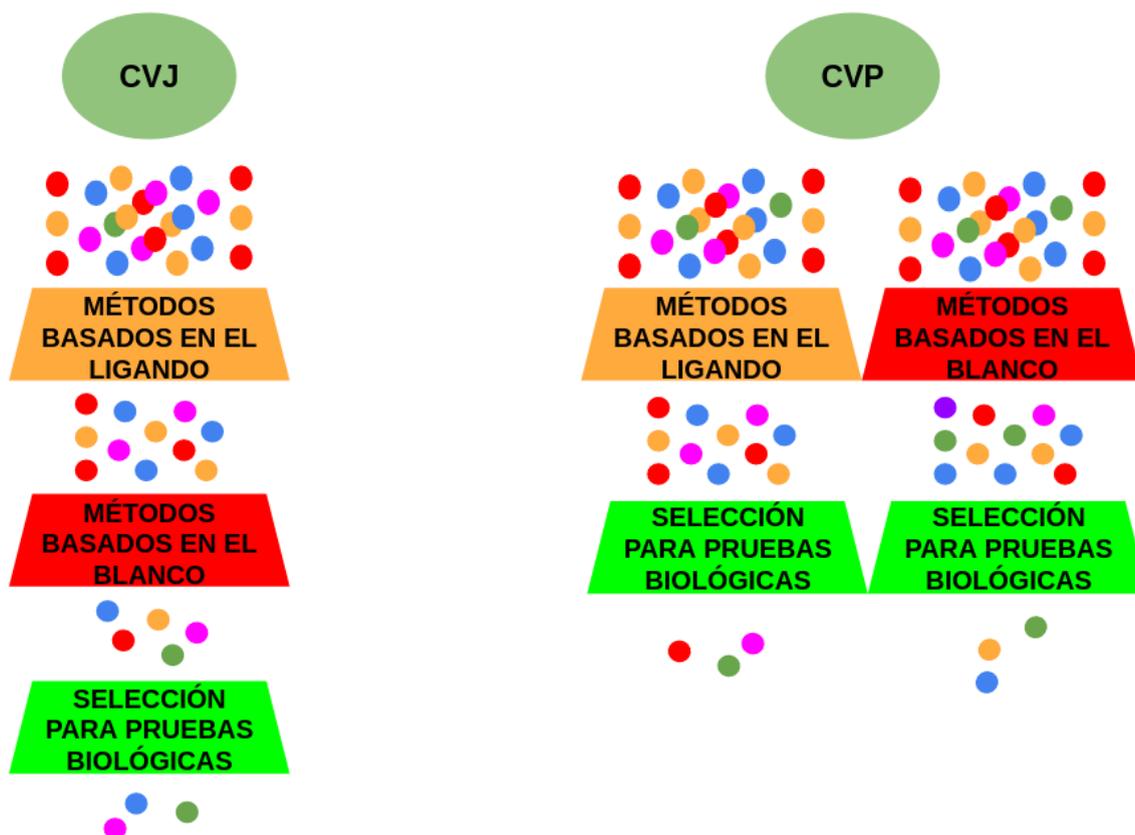


Figura 2.2.3. Diferentes estrategias combinadas para **CV**. Adaptado de Gavernet (2021).

2.3 REPOSICIONAMIENTO DE FÁRMACOS

2.3.1 GENERALIDADES

El reposicionamiento de fármacos, conocido en inglés como "*drug repositioning*" o "*drug repurposing*", se refiere a la búsqueda de nuevas aplicaciones terapéuticas para fármacos ya existentes. Los candidatos a ser reposicionados suelen ser fármacos que han sido aprobados previamente, retirados del mercado, abandonados en pruebas clínicas debido a su falta de eficacia para ciertas patologías, o aquellos que se encuentran en etapas avanzadas de investigación (Liu et al., 2013, Naveja et al., 2016).

La polifarmacología, la similitud entre blancos moleculares ortólogos en distintas especies y las conexiones existentes entre vías metabólicas son factores que permiten el uso de fármacos para múltiples indicaciones. Sin embargo, es importante tener en cuenta que una excesiva promiscuidad puede estar asociada a efectos adversos (Ma et al., 2013).

El reposicionamiento de fármacos es una estrategia prometedora en el campo farmacéutico ya que reduce significativamente los tiempos necesarios para desarrollar nuevas terapias y aumenta la probabilidad de éxito en los estudios clínicos (Temesi et al., 2014). Además, puede evitar la necesidad de etapas de deconvolución/identificación del blanco molecular, optimización del líder y pruebas de seguridad temprana, lo que se traduce en una reducción de los costos asociados (Swinney, 2013; Paul & Lewis-Hall, 2013; Rudrapal et al., 2020). Por esta razón, la industria farmacéutica, organismos gubernamentales y el sector académico han incrementado su interés en el reposicionamiento de fármacos (Liu et al., 2013; Polamreddy & Gattu, 2019), situación que se visibiliza, por ejemplo, en las ganancias del sector farmacéutico asociadas a reposicionamientos, las cuales fueron estimadas en 24400 millones de dólares estadounidenses en 2015 y proyectadas a 31300 millones en 2020 (Rudrapal et al., 2020).

La **Figura 2.3.1** muestra el aumento en la frecuencia de aparición de los términos "*drug repositioning*" y "*drug repurposing*" en los títulos y resúmenes de trabajos científicos indexados en Pubmed durante el período 2014-2023.

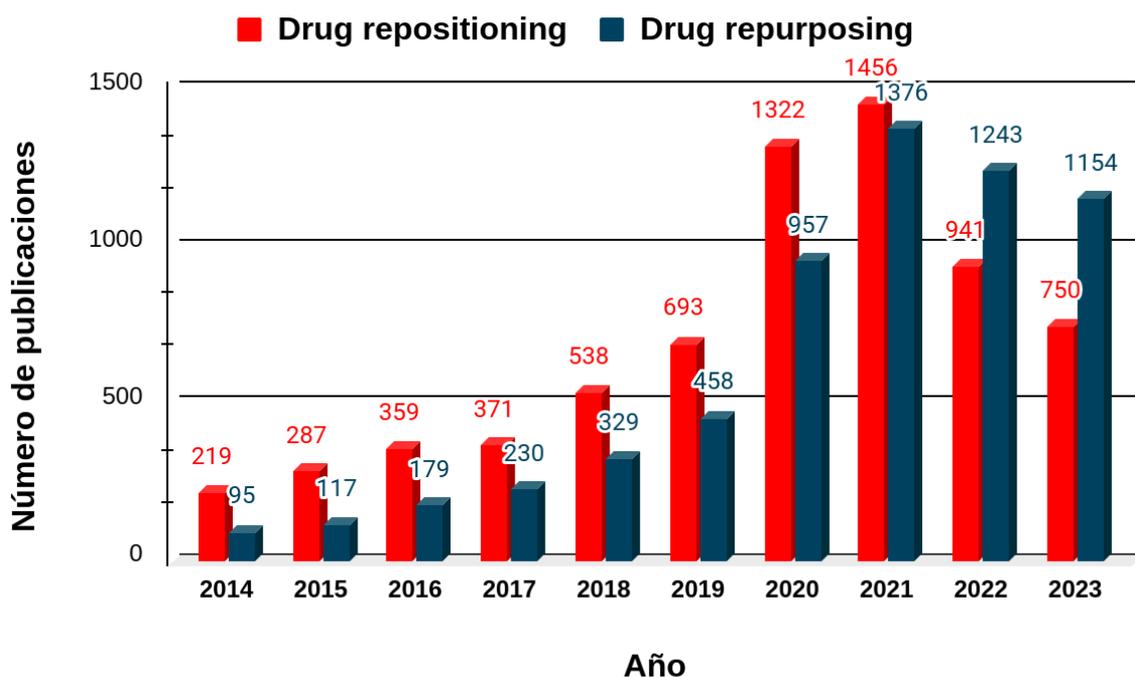


Figura 2.3.1. Número de artículos publicados en revistas científicas indexadas por Pubmed que contienen el término "*drug repositioning*" o "*drug repurposing*" en su título y/o resumen entre 2014 y 2023.

El reposicionamiento de fármacos ha dado lugar a casos emblemáticos como el del Sildenafil y la Talidomida. El Sildenafil se investigó inicialmente como antihipertensivo y para el tratamiento de la angina de pecho, pero se descubrió su efecto sobre la erección peneana durante las pruebas clínicas, lo que llevó a su

aprobación para el tratamiento de la disfunción eréctil (Ashburn & Thor, 2004; Barnett & Machado, 2006). La Talidomida, que se utilizó como antinauseoso en mujeres embarazadas en la década del '60 y se retiró del mercado debido a sus efectos teratogénicos, se redireccionó décadas después como tratamiento para las complicaciones cutáneas de la lepra denominadas **ENL** (eritemas nudosos leprosos). A pesar de su teratogenicidad, la **FDA** aprobó su uso para este propósito y, posteriormente, para el tratamiento del mieloma múltiple debido a su capacidad para inhibir la angiogénesis (Teo et al., 2002; Zhou et al., 2013; Singhal et al., 1999).

Aunque algunos casos han surgido casualmente (es decir, por serendipia), existen métodos basados en el conocimiento para la búsqueda de nuevos reposicionamientos (Liu et al., 2013; Naveja et al., 2016).

2.3.2 MÉTODOS SISTEMÁTICOS DE REPOSICIONAMIENTO DE FÁRMACOS

La búsqueda de relaciones entre fármacos, proteínas y enfermedades para el reposicionamiento de fármacos es realizada por métodos sistemáticos, como se muestra en la **Figura 2.3.2**. Los métodos de reposicionamiento se dividen en dos grandes grupos según la utilización o no de computadoras. Los métodos no computacionales involucran experimentos *in vitro*, como por ejemplo el **HTS**, y también se basan en la literatura. Por otro lado, los métodos computacionales se dividen en tres grandes grupos basados en la información que utilizan: métodos bioinformáticos, quimioinformáticos y literatura de alto rendimiento.

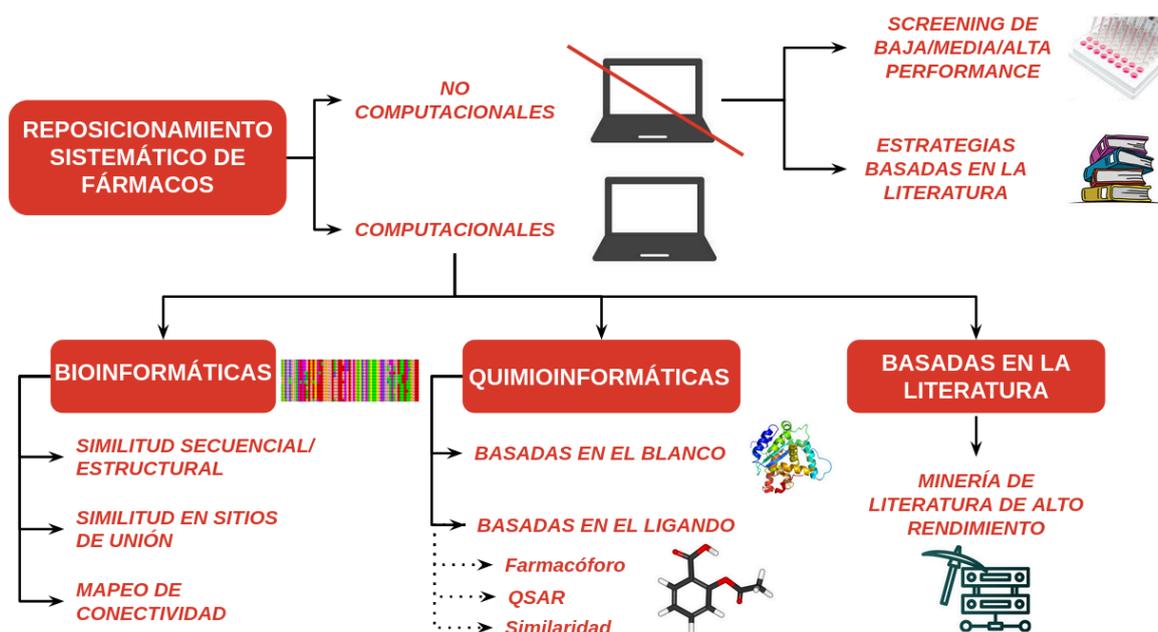


Figura 2.3.2. Esquema de los métodos sistemáticos que pueden ser aplicados en el reposicionamiento de fármacos. Adaptado de Talevi et al. (2018).

2.3.2.1 ENFOQUES NO COMPUTACIONALES

Descifrar el mecanismo de acción de un fármaco y comprender la fisiopatología de una enfermedad podría proporcionar supuestos sobre la posible utilización de un medicamento conocido para una nueva indicación. Estas aproximaciones son conocidas como estrategias de descubrimiento basadas en el conocimiento o la literatura (Naveja et al., 2016).

Entre las primeras estrategias sistemáticas empleadas para identificar posibles reposicionamientos también se encuentran las metodologías de tamizaje experimental. Estos métodos han evolucionado desde el tamizaje de baja capacidad (que examina sólo un pequeño conjunto de compuestos) hasta las estrategias de **HTS** ya mencionadas, que permiten evaluar más de 100.000 compuestos en solo un día (Hertzberg & Pope, 2000). Estos métodos representan una forma interesante de buscar nuevas oportunidades de reposicionamiento, ya que, además de evaluar el impacto en un objetivo molecular, hoy en día es posible adaptar los métodos al tamizaje fenotípico y, de esta manera, evaluar los medicamentos en sistemas complejos (Blay et al., 2020; Brito et al., 2020; Small et al., 2022).

Como se mencionó anteriormente, la principal limitación del **HTS** es el alto costo que requieren los equipos robóticos y la infraestructura, así como también el costo de disponer de la colección de compuestos y el alto número de falsos positivos detectados (Cronk, 2013).

2.3.2.2 ENFOQUES COMPUTACIONALES

Las técnicas computacionales brindan posibilidades muy valiosas para transformar el proceso fortuito de descubrir una nueva aplicación para un fármaco en una exploración sistemática y fundamentada de todas las opciones de reposicionamiento disponibles. El reposicionamiento *in silico* utiliza distintas estrategias para reunir, fusionar y analizar de forma sistemática grandes cantidades de datos, que incluyen bases de datos de fármacos previamente identificados (Liu et al., 2013; Prada Gori et al., 2023). Estas estrategias son de gran utilidad a la hora de organizar y explorar la información y conocimiento disponible para inferir nuevas asociaciones entre compuestos, blancos moleculares y enfermedades (Prada Gori et al., 2023).

2.3.2.2.1 METODOLOGÍAS QUIMIOINFORMÁTICAS

Existen diversos métodos de quimioinformática para el reposicionamiento de fármacos, pero los más comunes se basan en las técnicas de **CV** previamente discutidas. En la actualidad, hay varias bases de datos de acceso libre que contienen numerosos candidatos a fármacos reposicionados, tales como DrugBank (Wishart et al., 2018), Sweetlead (Novick et al., 2013) ó **DRH** (Corsello et al., 2017)

(del inglés *Drug Repurposing Hub*), así como información sobre dianas moleculares relacionadas con enfermedades humanas, como la *Therapeutic Target Database (TTD)* (Li et al., 2018).

2.3.2.2 METODOLOGÍAS BIOINFORMÁTICAS

Los avances en la genómica de los últimos años han llevado a la generación de grandes cantidades de datos genómicos y transcriptómicos de muestras de tejido patológico, muestras de tejido normal, modelos animales y líneas celulares. Gracias a la filosofía emergente de datos abiertos, muchos de estos datos se encuentran públicamente disponibles en diferentes repositorios públicamente accesibles (Talevi, 2023). Estos datos, combinados con las bases de datos que compilan información fenotípica y clínica, proporcionan una gran oportunidad para comprender mejor los mecanismos de las enfermedades, dilucidar los mecanismos de acción de moléculas bioactivas y descubrir nuevas indicaciones para fármacos ya conocidos. Uno de los tipos de datos más ampliamente utilizados para el reposicionamiento de fármacos son los perfiles transcriptómicos (Li et al., 2016).

Dos de las fuentes de información más importantes para utilizar en el reposicionamiento de fármacos son el proyecto *Connectivity Map* y su proyecto extendido **LINCS** (*Library of Integrated Network-Based Cellular Signatures*) (Lamb et al., 2006; Vidović et al., 2014). Ambos proyectos han producido perfiles de expresión génica a gran escala de tejidos y tipos celulares relevantes para una diversidad de enfermedades, incluyendo cáncer, enfermedades cardíacas y enfermedades neurodegenerativas. Estos perfiles de expresión génica han sido expuestos a diferentes agentes perturbadores lo que ha permitido construir un mapa detallado de las asociaciones funcionales entre enfermedades, perturbaciones genéticas y acciones de moléculas tipo fármaco. Esta base ha sido extensamente utilizada en estudios de reposicionamiento, especialmente en la estrategia conocida como "reversión de la firma" (en inglés *signature reversion*) (Li et al., 2016). Esta estrategia consiste en la identificación de moléculas capaces de revertir el perfil de expresión génica en líneas celulares patológicas. Por ejemplo, si se conoce el perfil de expresión génica causado por cada una de las moléculas en una línea celular dada, se puede seleccionar el perfil que sea más similar al perfil de expresión inverso de la línea celular patológica en estudio (Fisher et al., 2023).

Otra estrategia que se utiliza en el reposicionamiento de fármacos es la "culpable por asociación" (en inglés *guilt-by-association*), que busca moléculas que provocan respuestas transcripcionales similares, postulando que ellas deberían compartir similares mecanismos de acción (Iorio et al., 2013). Para llevar a cabo esta estrategia existen varios repositorios públicos disponibles como *Drug versus Disease (DvD)* (Pacini et al., 2013), *The Database for Annotation, Visualization and Integrated Discovery (DAVID)* (Dennis et al., 2003) y **MANTRA** (*Mode of Action by NeTwoRk Analysis*) (Carrella et al., 2014).

Además de estas estrategias, existen otras herramientas bioinformáticas que pueden ser utilizadas para el reposicionamiento de fármacos, como aquellas que comparan las secuencias genéticas o las estructuras de las proteínas. Estos métodos se basan en el postulado de que proteínas homólogas de diferentes especies pueden interactuar con los mismos ligandos (Gfeller & Zoete, 2015).

Finalmente, también se pueden lograr reposicionamientos mediante la comparativa de los sitios de unión. Proteínas con baja similitud de secuencia global, e incluso de distinta funcionalidad, pueden contar con sitios de unión similares desde el punto de vista estructural, lo cual les permitiría interactuar con ligandos similares. En consecuencia, es posible identificar mediante herramientas bioinformáticas sitios de unión similares en proteínas sin relación evolutiva, permitiendo encontrar posibles candidatos a reposicionamiento a partir de ligandos conocidos del sitio de unión similar al de interés (Konc et al., 2015; Haupt & Schroeder, 2011).

2.3.2.2.3 METODOLOGÍAS BASADAS EN LA LITERATURA

La estrategia denominada descubrimiento basado en la literatura (**DBL**), consiste en procesar datos de literatura para identificar relaciones implícitas o indirectas entre entidades biológicas que aparentemente no están conectadas entre sí (**Figura 2.3.2.2.3**). **DBL** analiza grandes volúmenes de literatura científica como un sistema de interconexiones entre términos de interés, y a través de ese análisis describe los mecanismos moleculares básicos que determinan una enfermedad, aumenta la comprensión de los efectos biológicos de una molécula, evalúa el perfil de riesgo/beneficio, y conduce a nuevos descubrimientos (Deftereos et al., 2011; Andronis et al., 2011).

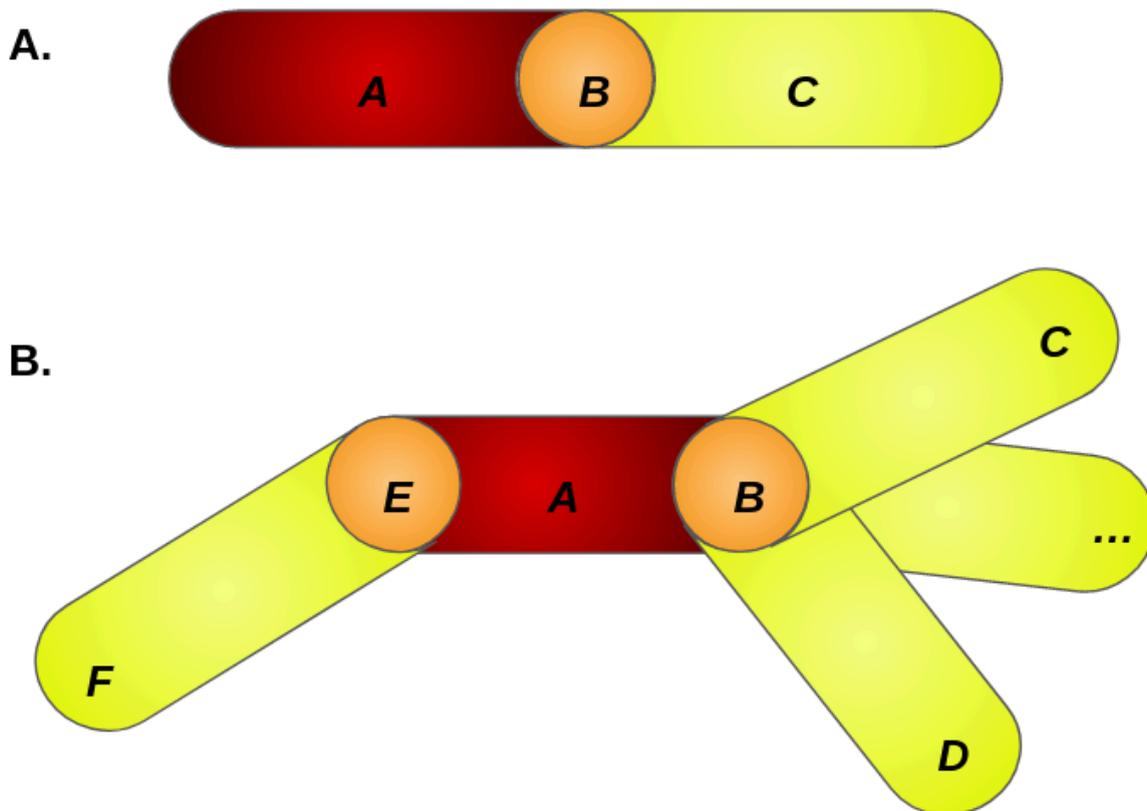


Figura 2.3.2.2.3. Modelos de metodologías basadas en la literatura. **(A)** El concepto A está relacionado al concepto B, el cual a su vez está relacionado con C como se informó en un set de trabajos científicos. A pesar de que no se conoce que A está directamente relacionado con C, se puede inferir una relación indirecta a través de B. **(B)** El concepto A está relacionado al concepto B, y a través de esto a C, D, etc. Además, A está relacionado a F a través de E. Adaptado de Deftereos et al. (2011).

2.3.3 REPOSICIONAMIENTO DE FÁRMACOS PARA ETDs

A pesar del alto costo humano y económico de las **ETDs**, el desarrollo de nuevas terapias para tratarlas es limitado en comparación con otras enfermedades. De hecho, de los 1393 fármacos aprobados entre 1975 y 1999, sólo el 0,93% estaba destinado a **ETDs**. En el período de 2000 a 2011, únicamente el 0,59% de los 850 productos terapéuticos registrados se destinaron a **ETDs**, y ninguno de ellos fue una entidad química novedosa. Más aún, solo el 1,65% de los productos farmacéuticos en la fase I de ensayos clínicos entre 2000 y 2014 se destinó al tratamiento o prevención de **ETDs** (Weng et al., 2018).

Debido a que gran parte del descubrimiento de fármacos para tratar las **ETDs** se lleva a cabo en laboratorios académicos sin los recursos y financiamiento de las compañías farmacéuticas, el reposicionamiento de fármacos se ha convertido en

una estrategia ampliamente adoptada para superar estas limitaciones (Klug et al., 2016).

Se han registrado diversos ejemplos de reposicionamiento de fármacos en el ámbito de las **ETDs**. Por ejemplo, la Eflornitina, inicialmente concebida como un agente anticancerígeno, se ha utilizado con éxito en el tratamiento de la Enfermedad del Sueño desde 1990 (Nwaka & Hudson, 2006). El Nifurtimox, un agente anti chagásico, también ha demostrado eficacia en el tratamiento de la Enfermedad del Sueño, en combinación con Eflornitina (Priotto et al., 2009; Kuemmerle et al., 2021).

Por otro lado, el Fexinidazol fue desarrollado como un agente antimicrobiano de amplio espectro en fase preclínica durante los años '80 (Torreele et al., 2010). Actualmente, este fármaco es la primera terapia oral disponible para la Enfermedad del Sueño en pacientes mayores a los 6 años de edad (Ferreira & Andricopulo, 2019). Otros ejemplos de fármacos reposicionados ya aprobados para el tratamiento de las **ETDs** incluyen a la Amiodarona, Anfotericina B, Itraconazol, Azitromicina y Miltefosina, entre otros (Sbaraglini et al., 2016; Dorlo et al., 2012).

2.3.3.1 CASOS EXITOSOS EN CESTODOS MEDIANTE ENFOQUES COMPUTACIONALES

En el capítulo 1 de la presente tesis doctoral se presentaron una serie de compuestos candidatos a reposicionamiento de fármacos tanto para la equinocosis y/o cisticercosis, en donde se hizo una descripción de cada uno de ellos. Sin embargo, la literatura es escasa en cuanto al reposicionamiento de fármacos por métodos computacionales aplicados al potencial tratamiento de enfermedades causadas por cestodos.

Bélgamo et al. (2020) publicaron un primer estudio en donde se lleva adelante un **CV** basado en modelos **QSAR** categóricos, utilizando como información 258 compuestos con actividad reportada contra **AFABP**. De dicho **CV**, los autores seleccionaron al Sulindac (un antiinflamatorio no esteroideo), Cetilpiridinio (antiséptico utilizado en productos de higiene personal) y Clotrimazol (un agente antimicótico) para las pruebas *in vitro*. Mediante ensayos de desplazamiento por fluorescencia utilizando al Ácido 8-anilino naftaleno-1-sulfónico (**ANS**) como molécula fluorescente y al Ácido oleico como ligando natural de las **FABPs**, determinaron las constantes aparentes de disociación para cada uno de los candidatos a reposicionamiento en tres isoformas de **FABPs** de *E. multilocularis*, expresadas de manera recombinante. Los autores observaron que todos los compuestos ensayados pudieron desplazar al **ANS** en al menos 1 isoforma de **FABP** a concentración 15 μM . De esta manera, dicho trabajo demostró, *in vitro*, la capacidad de los modelos **QSAR** para la predicción de compuestos candidatos a reposicionamiento, como así también el éxito de la implementación del paradigma de reposicionamiento de blancos.

El paradigma de reposicionamiento de blancos es una forma relativamente nueva de abordar sistemáticamente la búsqueda de nuevos tratamientos contra agentes infecciosos, incluyendo las **ETDs** (Klug et al., 2016). Muy brevemente, se comienza con un blanco parasitario definido que tiene un homólogo establecido en otra especie. Los fármacos que se dirigen a esta segunda especie se exploran como punto de partida para desarrollar compuestos que inhiban al agente infeccioso buscado. Los compuestos identificados mediante esta estrategia pueden ser luego sometidos a una optimización molecular para mejorar su selectividad al blanco molecular deseado. Se pueden mencionar otros ejemplos de éxito en la implementación de esta estrategia como el trabajo de Nallan y colaboradores en donde evaluaron la capacidad de un panel de inhibidores de la transferasa de farnesilo de proteínas de mamíferos en desarrollo preclínico y clínico para inhibir el crecimiento *in vitro* de parásitos de *Plasmodium falciparum* (Nallan et al., 2005). Del mismo modo, otros autores han utilizado inhibidores humanos de fosfodiesterasas como puntos de partida potenciales para el desarrollo de nuevos tratamientos contra la tripanosomiasis africana (Bland et al., 2011; Amata et al., 2014), dirigidos a las fosfodiesterasas de *Trypanosoma brucei*.

2.4 ESTRATEGIA UTILIZADA EN ESTE TRABAJO DE TESIS

En la presente tesis doctoral se han desarrollado modelos computacionales para descubrir inhibidores que pudieran afectar la incorporación de lípidos en los parásitos *E. granulosus*, *E. multilocularis* y *T. solium*, a partir de bases de datos de compuestos químicos. Para llevar adelante el **CV**, se ha hecho uso de la estrategia de **CVP**, utilizando modelos **QSAR-2D** basados en descriptores moleculares obtenidos a través de métodos indirectos (basado en ligandos), como así también modelos de *docking* molecular utilizando las estructuras tridimensionales de las distintas isoformas de las **FABPs** de dichos parásitos, la gran mayoría de ellas obtenidas mediante modelado por aprendizaje automático y **DM**.

El **QSAR-2D** es una técnica rápida y económica debido a que no es necesaria la optimización previa de las conformaciones moleculares, pero cuenta como principal desventaja a la dificultad que tienen estos modelos a la hora de la interpretación de los resultados, es decir, la relación entre actividad y estructura no es sencilla de inferir. Por otro lado, los métodos directos de cribado permiten una interpretación directa del fenómeno de unión entre el receptor y su ligando, como así también una adecuada discriminación de actividad entre isómeros geométricos. Sin embargo, la etapa de optimización de las bibliotecas químicas, como así también la etapa de cálculo por parte de los algoritmos de dinámica y *docking* molecular son costosas computacionalmente, teniendo requerimientos de hardware (ya sea tarjetas gráficas

o memoria **RAM** adecuadas) que posiblemente escapen a un computador ordinario de escritorio.

Debido al éxito de las estrategias de reposicionamiento de fármacos para la búsqueda de nuevos tratamientos para **ETDs**, las ventajas de las diferentes técnicas de cribado *in silico*, y la complementariedad de las mismas, ambos enfoques han sido aplicados en el **CV** de bases de datos de compuestos químicos.

REFERENCIAS

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>

Alberca, Lucas N. (2018). *Búsqueda asistida por computadora de nuevos fármacos antichagásicos análogos de poliaminas* [Doctor en Ciencias Exactas, área Ciencias Biológicas, Universidad Nacional de La Plata]. <https://doi.org/10.35537/10915/72607>

Alberts, I. L., Todorov, N. P., & Dean, P. M. (2005). Receptor Flexibility in de Novo Ligand Design and Docking. *Journal of Medicinal Chemistry*, 48(21), 6585–6596. <https://doi.org/10.1021/jm050196j>

Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., & Smith, J. C. (2018). Ensemble Docking in Drug Discovery. *Biophysical Journal*, 114(10), 2271–2278. <https://doi.org/10.1016/j.bpj.2018.02.038>

Amata, E., Bland, N. D., Hoyt, C. T., Settimo, L., Campbell, R. K., & Pollastri, M. P. (2014). Repurposing human PDE4 inhibitors for neglected tropical diseases: Design, synthesis and evaluation of cilomilast analogues as *Trypanosoma brucei* PDEB1 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 24(17), 4084–4089. <https://doi.org/10.1016/j.bmcl.2014.07.063>

Andrade, C. H., Pasqualoto, K. F. M., Ferreira, E. I., & Hopfinger, A. J. (2010). 4D-QSAR: Perspectives in Drug Design. *Molecules*, 15(5), 3281–3294. <https://doi.org/10.3390/molecules15053281>

Andronis, C., Sharma, A., Virvilis, V., Deftereos, S., & Persidis, A. (2011). Literature mining, ontologies and information visualization for drug repurposing. *Briefings in Bioinformatics*, 12(4), 357–368. <https://doi.org/10.1093/bib/bbr005>

Andrusier, N., Mashiach, E., Nussinov, R., & Wolfson, H. J. (2008). Principles of flexible protein-protein docking. *Proteins: Structure, Function, and Bioinformatics*, 73(2), 271–289. <https://doi.org/10.1002/prot.22170>

Ashburn, T. T., & Thor, K. B. (2004). Drug repositioning: Identifying and developing new uses for existing drugs. *Nature Reviews Drug Discovery*, 3(8), 673–683. <https://doi.org/10.1038/nrd1468>

Bajorath, J. (2001). Selected Concepts and Investigations in Compound Classification, Molecular Descriptor Analysis, and Virtual Screening. *Journal of*

Chemical Information and Computer Sciences, 41(2), 233–245.
<https://doi.org/10.1021/ci0001482>

Ball, Philip. (2020). *Behind the screens of AlphaFold*. Chemistry World.
<https://www.chemistryworld.com/opinion/behind-the-screens-of-alphafold/4012867.article>

Barnett, C. F., & Machado, R. F. (2006). Sildenafil in the treatment of pulmonary hypertension. *Vascular Health and Risk Management*, 2(4), 411–422.
<https://doi.org/10.2147/vhrm.2006.2.4.411>

Bélgamo, J. A., Alberca, L. N., Pórfido, J. L., Romero, F. N. C., Rodriguez, S., Talevi, A., Córscico, B., & Franchini, G. R. (2020). Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets. *Journal of Computer-Aided Molecular Design*, 34(12), 1275–1288. <https://doi.org/10.1007/s10822-020-00352-8>

Berendsen, H. J. C., Van Der Spoel, D., & Van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1–3), 43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)

Bland, N. D., Wang, C., Tallman, C., Gustafson, A. E., Wang, Z., Ashton, T. D., Ochiana, S. O., McAllister, G., Cotter, K., Fang, A. P., Gechijian, L., Garceau, N., Gangurde, R., Ortenberg, R., Ondrechen, M. J., Campbell, R. K., & Pollastri, M. P. (2011). Pharmacological Validation of *Trypanosoma brucei* Phosphodiesterases B1 and B2 as Druggable Targets for African Sleeping Sickness. *Journal of Medicinal Chemistry*, 54(23), 8188–8194. <https://doi.org/10.1021/jm201148s>

Blay, V., Tolani, B., Ho, S. P., & Arkin, M. R. (2020). High-Throughput Screening: Today's biochemical and cell-based approaches. *Drug Discovery Today*, 25(10), 1807–1821. <https://doi.org/10.1016/j.drudis.2020.07.024>

Brito, H., Marques, V., Afonso, M. B., Brown, D. G., Börjesson, U., Selmi, N., Smith, D. M., Roberts, I. O., Fitzek, M., Aniceto, N., Guedes, R. C., Moreira, R., & Rodrigues, C. M. P. (2020). Phenotypic high-throughput screening platform identifies novel chemotypes for necroptosis inhibition. *Cell Death Discovery*, 6(1), 6. <https://doi.org/10.1038/s41420-020-0240-0>

Carrella, D., Napolitano, F., Rispoli, R., Miglietta, M., Carissimo, A., Cutillo, L., Sirci, F., Gregoretti, F., & Di Bernardo, D. (2014). Mantra 2.0: An online collaborative resource for drug mode of action and repurposing by network analysis. *Bioinformatics*, 30(12), 1787–1788. <https://doi.org/10.1093/bioinformatics/btu058>

Carroni, M., & Saibil, H. R. (2016). Cryo electron microscopy to determine the structure of macromolecular complexes. *Methods*, 95, 78–85. <https://doi.org/10.1016/j.ymeth.2015.11.023>

Consonni, V., & Todeschini, R. (2012). Multivariate Analysis of Molecular Descriptors. In M. Dehmer, K. Varmuza, & D. Bonchev (Eds.), *Statistical Modelling of Molecular Descriptors in QSAR/QSPR* (1st ed., pp. 111–147). Wiley. <https://doi.org/10.1002/9783527645121.ch4>

Cook, D., Brown, D., Alexander, R., March, R., Morgan, P., Satterthwaite, G., & Pangalos, M. N. (2014). Lessons learned from the fate of AstraZeneca's drug pipeline: A five-dimensional framework. *Nature Reviews Drug Discovery*, 13(6), 419–431. <https://doi.org/10.1038/nrd4309>

Corsello, S. M., Bittker, J. A., Liu, Z., Gould, J., McCarren, P., Hirschman, J. E., Johnston, S. E., Vrcic, A., Wong, B., Khan, M., Asiedu, J., Narayan, R., Mader, C. C., Subramanian, A., & Golub, T. R. (2017). The Drug Repurposing Hub: A next-generation drug library and information resource. *Nature Medicine*, 23(4), 405–408. <https://doi.org/10.1038/nm.4306>

Council for International Organization of Medical Sciences & The International Council for Laboratory Animal Science (2012). International Guiding Principles for Biomedical Research Involving Animals.

Cronk, D. (2013). High-throughput screening. In *Drug Discovery and Development* (pp. 95–117). Elsevier. <https://doi.org/10.1016/B978-0-7020-4299-7.00008-1>

Curry, Stephen (2020). No, DeepMind has not solved protein folding. *Reciprocal Space*. <https://occamstypewriter.org/scurry/2020/12/02/no-deepmind-has-not-solved-protein-folding/>

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., & Overington, J. P. (2015). ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Research*, 43(W1), W612–W620. <https://doi.org/10.1093/nar/gkv352>

Davis & Ward. (2014). *The Handbook of Medicinal Chemistry: Principles and Practice*. Royal Society of Chemistry. <https://books.google.com.ar/books?id=mICsEAAAQBAJ>

De, P., Kar, S., Ambure, P., & Roy, K. (2022). Prediction reliability of QSAR models: An overview of various validation tools. *Archives of Toxicology*, 96(5), 1279–1295. <https://doi.org/10.1007/s00204-022-03252-y>

Dennis, G., Sherman, B. T., Hosack, D. A., Yang, J., Gao, W., Lane, H. C., & Lempicki, R. A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome Biology*, 4(9), R60. <https://doi.org/10.1186/gb-2003-4-9-r60>

Dorlo, T. P. C., Balasegaram, M., Beijnen, J. H., & De Vries, P. J. (2012). Miltefosine: A review of its pharmacology and therapeutic efficacy in the treatment of leishmaniasis. *Journal of Antimicrobial Chemotherapy*, 67(11), 2576–2597. <https://doi.org/10.1093/jac/dks275>

Douglas Heaven, Will. (2020). *DeepMind's protein-folding AI has solved a 50-year-old grand challenge of biology*. MIT Technology Review. <https://www.technologyreview.com/2020/11/30/1012712/deepmind-protein-folding-ai-solved-biology-science-drugs-disease/>

Eder, J., Sedrani, R., & Wiesmann, C. (2014). The discovery of first-in-class drugs: Origins and evolution. *Nature Reviews Drug Discovery*, 13(8), 577–587. <https://doi.org/10.1038/nrd4336>

Fernández-Torras, A., Comajuncosa-Creus, A., Duran-Frigola, M., & Aloy, P. (2022). Connecting chemistry and biology through molecular descriptors. *Current Opinion in Chemical Biology*, 66, 102090. <https://doi.org/10.1016/j.cbpa.2021.09.001>

Ferrari, A. M., Wei, B. Q., Costantino, L., & Shoichet, B. K. (2004). Soft Docking and Multiple Receptor Conformations in Virtual Screening. *Journal of Medicinal Chemistry*, 47(21), 5076–5084. <https://doi.org/10.1021/jm049756p>

Ferreira, L. L. G., & Andricopulo, A. D. (2019). Drugs and vaccines in the 21st century for neglected diseases. *The Lancet Infectious Diseases*, 19(2), 125–127. [https://doi.org/10.1016/S1473-3099\(19\)30005-2](https://doi.org/10.1016/S1473-3099(19)30005-2)

Ferreira, L. L. G., De Moraes, J., & Andricopulo, A. D. (2022). Approaches to advance drug discovery for neglected tropical diseases. *Drug Discovery Today*, 27(8), 2278–2287. <https://doi.org/10.1016/j.drudis.2022.04.004>

Fiser, A. (2010). Template-Based Protein Structure Modeling. In D. Fenyö (Ed.), *Computational Biology* (Vol. 673, pp. 73–94). Humana Press. https://doi.org/10.1007/978-1-60761-842-3_6

Fisher, J. L., Wilk, E. J., Oza, V. H., Howton, T. C., Flanary, V., Clark, A. D., Hjelmeland, A. B., & Lasseigne, B. N. (2023). *Signature reversion of three disease-associated gene signatures prioritizes cancer drug repurposing candidates* [Preprint]. Genomics. <https://doi.org/10.1101/2023.03.10.532074>

Gaitonde, V., Karmakar, P., & Trivedi, A. (Eds.). (2020). *Drug Discovery and Development—New Advances*. IntechOpen. <https://doi.org/10.5772/intechopen.77685>

Gaulton, A., Hersey, A., Nowotka, M., Bento, A. P., Chambers, J., Mendez, D., Mutowo, P., Atkinson, F., Bellis, L. J., Cibrián-Uhalte, E., Davies, M., Dedman, N., Karlsson, A., Magariños, M. P., Overington, J. P., Papadatos, G., Smit, I., & Leach, A. R. (2017). The ChEMBL database in 2017. *Nucleic Acids Research*, 45(D1), D945–D954. <https://doi.org/10.1093/nar/gkw1074>

Gavernet, L. (Ed.). (2021). *Introducción a la Química Medicinal*. Editorial de la Universidad Nacional de La Plata (EDULP). <https://doi.org/10.35537/10915/114312>

Gfeller, D., & Zoete, V. (2015). Protein homology reveals new targets for bioactive small molecules. *Bioinformatics*, 31(16), 2721–2727. <https://doi.org/10.1093/bioinformatics/btv214>

González-Díaz, H., Prado-Prado, F., García-Mera, X., Alonso, N., Abeijón, P., Caamaño, O., Yáñez, M., Munteanu, C. R., Pazos, A., Dea-Ayuela, M. A., Gómez-Muñoz, M. T., Garijo, M. M., Sansano, J., & Ubeira, F. M. (2011).

MIND-BEST: Web Server for Drugs and Target Discovery; Design, Synthesis, and Assay of MAO-B Inhibitors and Theoretical–Experimental Study of G3PDH Protein from *Trichomonas gallinae*. *Journal of Proteome Research*, 10(4), 1698–1718. <https://doi.org/10.1021/pr101009e>

Grisoni, F., Ballabio, D., Todeschini, R., & Consonni, V. (2018). Molecular Descriptors for Structure–Activity Applications: A Hands-On Approach. In O. Nicolotti (Ed.), *Computational Toxicology* (Vol. 1800, pp. 3–53). Springer New York. https://doi.org/10.1007/978-1-4939-7899-1_1

Guerry, P., & Herrmann, T. (2011). Advances in automated NMR protein structure determination. *Quarterly Reviews of Biophysics*, 44(3), 257–309. <https://doi.org/10.1017/S0033583510000326>

Haupt, V. J., & Schroeder, M. (2011). Old friends in new guise: Repositioning of known drugs with structural bioinformatics. *Briefings in Bioinformatics*, 12(4), 312–326. <https://doi.org/10.1093/bib/bbr011>

Hertzberg, R. P., & Pope, A. J. (2000). High-throughput screening: New technology for the 21st century. *Current Opinion in Chemical Biology*, 4(4), 445–451. [https://doi.org/10.1016/S1367-5931\(00\)00110-1](https://doi.org/10.1016/S1367-5931(00)00110-1)

Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3), 435–447. <https://doi.org/10.1021/ct700301q>

Hollingsworth, S. A., & Dror, R. O. (2018). Molecular Dynamics Simulation for All. *Neuron*, 99(6), 1129–1143. <https://doi.org/10.1016/j.neuron.2018.08.011>

Hughes, J., Rees, S., Kalindjian, S., & Philpott, K. (2011). Principles of early drug discovery: Principles of early drug discovery. *British Journal of Pharmacology*, 162(6), 1239–1249. <https://doi.org/10.1111/j.1476-5381.2010.01127.x>

Ianni, M., Gantner, M., Ruiz, M., Castro, E., Bruno-Blanch, L., & Talevi, A. (2015). Systematic Comparison of the Performance of Different 2D and 3D Ligand-Based Virtual Screening Methodologies to Discover Anticonvulsant Drugs. *Combinatorial Chemistry & High Throughput Screening*, 18(4), 387–398. <https://doi.org/10.2174/1386207318666150305151420>

Ilari, A., & Savino, C. (2008). Protein Structure Determination by X-Ray Crystallography. In J. M. Keith (Ed.), *Bioinformatics* (Vol. 452, pp. 63–87). Humana Press. https://doi.org/10.1007/978-1-60327-159-2_3

Iorio, F., Rittman, T., Ge, H., Menden, M., & Saez-Rodriguez, J. (2013). Transcriptional data: A new gateway to drug repositioning? *Drug Discovery Today*, 18(7–8), 350–357. <https://doi.org/10.1016/j.drudis.2012.07.014>

Irwin, J. J., Sterling, T., Mysinger, M. M., Bolstad, E. S., & Coleman, R. G. (2012). ZINC: A Free Tool to Discover Chemistry for Biology. *Journal of Chemical Information and Modeling*, 52(7), 1757–1768. <https://doi.org/10.1021/ci3001277>

Irwin, J. J., Tang, K. G., Young, J., Dandarchuluun, C., Wong, B. R., Khurelbaatar, M., Moroz, Y. S., Mayfield, J., & Sayle, R. A. (2020). ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *Journal of Chemical Information and Modeling*, 60(12), 6065–6073. <https://doi.org/10.1021/acs.jcim.0c00675>

Jorgensen, W. L., Maxwell, D. S., & Tirado-Rives, J. (1996). Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118(45), 11225–11236. <https://doi.org/10.1021/ja9621760>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R.,

Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature*, 596(7873), 583–589. <https://doi.org/10.1038/s41586-021-03819-2>

Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, 9(9), 646–652. <https://doi.org/10.1038/nsb0902-646>

Kaserer, T., Beck, K., Akram, M., Odermatt, A., & Schuster, D. (2015). Pharmacophore Models and Pharmacophore-Based Virtual Screening: Concepts and Applications Exemplified on Hydroxysteroid Dehydrogenases. *Molecules*, 20(12), 22799–22832. <https://doi.org/10.3390/molecules201219880>

Katiyar, R. S., & Jha, P. K. (2018). Molecular simulations in drug delivery: Opportunities and challenges. *WIREs Computational Molecular Science*, 8(4), e1358. <https://doi.org/10.1002/wcms.1358>

Kim, S., Thiessen, P. A., Bolton, E. E., Chen, J., Fu, G., Gindulyte, A., Han, L., He, J., He, S., Shoemaker, B. A., Wang, J., Yu, B., Zhang, J., & Bryant, S. H. (2016). PubChem Substance and Compound databases. *Nucleic Acids Research*, 44(D1), D1202–D1213. <https://doi.org/10.1093/nar/gkv951>

Kirchmair, J., Markt, P., Distinto, S., Wolber, G., & Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *Journal of Computer-Aided Molecular Design*, 22(3–4), 213–228. <https://doi.org/10.1007/s10822-007-9163-6>

Kiriiri, G. K., Njogu, P. M., & Mwangi, A. N. (2020). Exploring different approaches to improve the success of drug discovery and development projects: A review. *Future Journal of Pharmaceutical Sciences*, 6(1), 27. <https://doi.org/10.1186/s43094-020-00047-9>

Klug, D. M., Gelb, M. H., & Pollastri, M. P. (2016). Repurposing strategies for tropical disease drug discovery. *Bioorganic & Medicinal Chemistry Letters*, 26(11), 2569–2576. <https://doi.org/10.1016/j.bmcl.2016.03.103>

Konc, J., Miller, B. T., Štular, T., Lešnik, S., Woodcock, H. L., Brooks, B. R., & Janežič, D. (2015). ProBiS-CHARMMing: Web Interface for Prediction and Optimization of Ligands in Protein Binding Sites. *Journal of Chemical Information and Modeling*, 55(11), 2308–2314. <https://doi.org/10.1021/acs.jcim.5b00534>

Kuemmerle, A., Schmid, C., Bernhard, S., Kande, V., Mutombo, W., Ilunga, M., Lumpungu, I., Mutanda, S., Nganzobo, P., Tete, D. N., Kisala, M., Burri, C., Blesson,

S., & Valverde Mordt, O. (2021). Effectiveness of Nifurtimox Eflornithine Combination Therapy (NECT) in *T. b. gambiense* second stage sleeping sickness patients in the Democratic Republic of Congo: Report from a field study. *PLOS Neglected Tropical Diseases*, *15*(11), e0009903. <https://doi.org/10.1371/journal.pntd.0009903>

Kuhlman, B., & Bradley, P. (2019). Advances in protein structure prediction and design. *Nature Reviews Molecular Cell Biology*, *20*(11), 681–697. <https://doi.org/10.1038/s41580-019-0163-x>

Kumar, A., & Zhang, K. Y. J. (2015). Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*, *71*, 26–37. <https://doi.org/10.1016/j.jymeth.2014.07.007>

Lamb, J., Crawford, E. D., Peck, D., Modell, J. W., Blat, I. C., Wrobel, M. J., Lerner, J., Brunet, J.-P., Subramanian, A., Ross, K. N., Reich, M., Hieronymus, H., Wei, G., Armstrong, S. A., Haggarty, S. J., Clemons, P. A., Wei, R., Carr, S. A., Lander, E. S., & Golub, T. R. (2006). The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science*, *313*(5795), 1929–1935. <https://doi.org/10.1126/science.1132939>

Law, V., Knox, C., Djoumbou, Y., Jewison, T., Guo, A. C., Liu, Y., Maciejewski, A., Arndt, D., Wilson, M., Neveu, V., Tang, A., Gabriel, G., Ly, C., Adamjee, S., Dame, Z. T., Han, B., Zhou, Y., & Wishart, D. S. (2014). DrugBank 4.0: Shedding new light on drug metabolism. *Nucleic Acids Research*, *42*(D1), D1091–D1097. <https://doi.org/10.1093/nar/gkt1068>

Lexa, K. W., & Carlson, H. A. (2012). Protein flexibility in docking and surface mapping. *Quarterly Reviews of Biophysics*, *45*(3), 301–343. <https://doi.org/10.1017/S0033583512000066>

Li, J., Zheng, S., Chen, B., Butte, A. J., Swamidass, S. J., & Lu, Z. (2016). A survey of current trends in computational drug repositioning. *Briefings in Bioinformatics*, *17*(1), 2–12. <https://doi.org/10.1093/bib/bbv020>

Li, Y. H., Yu, C. Y., Li, X. X., Zhang, P., Tang, J., Yang, Q., Fu, T., Zhang, X., Cui, X., Tu, G., Zhang, Y., Li, S., Yang, F., Sun, Q., Qin, C., Zeng, X., Chen, Z., Chen, Y. Z., & Zhu, F. (2018). Therapeutic target database update 2018: Enriched resource for facilitating bench-to-clinic research of targeted therapeutics. *Nucleic Acids Research*, *46*(D1), D1121–D1127. <https://doi.org/10.1093/nar/gkx1076>

Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, *7*(8), 306–317. <https://doi.org/10.1007/s008940100045>

Liu, X., Ouyang, S., Yu, B., Liu, Y., Huang, K., Gong, J., Zheng, S., Li, Z., Li, H., & Jiang, H. (2010). PharmMapper server: A web server for potential drug target identification using pharmacophore mapping approach. *Nucleic Acids Research*, 38(suppl_2), W609–W614. <https://doi.org/10.1093/nar/gkq300>

Liu, Z., Fang, H., Reagan, K., Xu, X., Mendrick, D. L., Slikker, W., & Tong, W. (2013). In silico drug repositioning – what we need to know. *Drug Discovery Today*, 18(3–4), 110–115. <https://doi.org/10.1016/j.drudis.2012.08.005>

Llanos, M. A., Sbaraglini, M. L., Villalba, M. L., Ruiz, M. D., Carrillo, C., Alba Soto, C., Talevi, A., Angeli, A., Parkkila, S., Supuran, C. T., & Gavernet, L. (2020). A structure-based approach towards the identification of novel antichagasic compounds: *Trypanosoma cruzi* carbonic anhydrase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 35(1), 21–30. <https://doi.org/10.1080/14756366.2019.1677638>

Luo, J., Wei, W., Waldispühl, J., & Moitessier, N. (2019). Challenges and current status of computational methods for docking small molecules to nucleic acids. *European Journal of Medicinal Chemistry*, 168, 414–425. <https://doi.org/10.1016/j.ejmech.2019.02.046>

Ma, D.-L., Chan, D. S.-H., & Leung, C.-H. (2013). Drug repositioning by structure-based virtual screening. *Chemical Society Reviews*, 42(5), 2130. <https://doi.org/10.1039/c2cs35357a>

Macalino, S. J. Y., Gosu, V., Hong, S., & Choi, S. (2015). Role of computer-aided drug design in modern drug discovery. *Archives of Pharmacal Research*, 38(9), 1686–1701. <https://doi.org/10.1007/s12272-015-0640-5>

Maggiore, G., Vogt, M., Stumpfe, D., & Bajorath, J. (2014). Molecular Similarity in Medicinal Chemistry. *Journal of Medicinal Chemistry*, 57(8), 3186–3204. <https://doi.org/10.1021/jm401411z>

Meng, X.-Y., Zhang, H.-X., Mezei, M., & Cui, M. (2011). Molecular Docking: A Powerful Approach for Structure-Based Drug Discovery. *Current Computer Aided-Drug Design*, 7(2), 146–157. <https://doi.org/10.2174/157340911795677602>

Mignani, S., Huber, S., Tomás, H., Rodrigues, J., & Majoral, J.-P. (2016). Why and how have drug discovery strategies in pharma changed? What are the new mindsets? *Drug Discovery Today*, 21(2), 239–249. <https://doi.org/10.1016/j.drudis.2015.09.007>

Miles, J. A., & Ross, B. P. (2021). Recent Advances in Virtual Screening for Cholinesterase Inhibitors. *ACS Chemical Neuroscience*, 12(1), 30–41. <https://doi.org/10.1021/acscemneuro.0c00627>

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, 19(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>

Moffat, J. G., Vincent, F., Lee, J. A., Eder, J., & Prunotto, M. (2017). Opportunities and challenges in phenotypic drug discovery: An industry perspective. *Nature Reviews Drug Discovery*, 16(8), 531–543. <https://doi.org/10.1038/nrd.2017.111>

Mohs, R. C., & Greig, N. H. (2017). Drug discovery and development: Role of basic biological research. *Alzheimer's & Dementia: Translational Research & Clinical Interventions*, 3(4), 651–657. <https://doi.org/10.1016/j.trci.2017.10.005>

Moriwaki, H., Tian, Y.-S., Kawashita, N., & Takagi, T. (2018). Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, 10(1), 4. <https://doi.org/10.1186/s13321-018-0258-y>

Muhammed, M. T., & Aki-Yalcin, E. (2021). Pharmacophore Modeling in Drug Discovery: Methodology and Current Status. *Journal of the Turkish Chemical Society Section A: Chemistry*, 8(3), 749–762. <https://doi.org/10.18596/jotcsa.927426>

Nallan, L., Bauer, K. D., Bendale, P., Rivas, K., Yokoyama, K., Hornéy, C. P., Pendyala, P. R., Floyd, D., Lombardo, L. J., Williams, D. K., Hamilton, A., Sebti, S., Windsor, W. T., Weber, P. C., Buckner, F. S., Chakrabarti, D., Gelb, M. H., & Van Voorhis, W. C. (2005). Protein Farnesyltransferase Inhibitors Exhibit Potent Antimalarial Activity. *Journal of Medicinal Chemistry*, 48(11), 3704–3713. <https://doi.org/10.1021/jm0491039>

National Institutes of Health (NIH). (2024). *PubMed*. PubMed. Retrieved February 7, 2024, from <https://pubmed.ncbi.nlm.nih.gov/>

Naveja, J. J., Dueñas-González, A., & Medina-Franco, J. L. (2016). Drug Repurposing for Epigenetic Targets Guided by Computational Methods. In *Epi-Informatics* (pp. 327–357). Elsevier. <https://doi.org/10.1016/B978-0-12-802808-7.00012-5>

Neves, B. J., Braga, R. C., Melo-Filho, C. C., Moreira-Filho, J. T., Muratov, E. N., & Andrade, C. H. (2018). QSAR-Based Virtual Screening: Advances and Applications in Drug Discovery. *Frontiers in Pharmacology*, 9, 1275. <https://doi.org/10.3389/fphar.2018.01275>

Nogales, E., & Scheres, S. H. W. (2015). Cryo-EM: A Unique Tool for the Visualization of Macromolecular Complexity. *Molecular Cell*, 58(4), 677–689. <https://doi.org/10.1016/j.molcel.2015.02.019>

Novick, P. A., Ortiz, O. F., Poelman, J., Abdulhay, A. Y., & Pande, V. S. (2013). SWEETLEAD: An In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PLoS ONE*, 8(11), e79568. <https://doi.org/10.1371/journal.pone.0079568>

Nwaka, S., & Hudson, A. (2006). Innovative lead discovery strategies for tropical diseases. *Nature Reviews Drug Discovery*, 5(11), 941–955. <https://doi.org/10.1038/nrd2144>

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, 3(1), 33. <https://doi.org/10.1186/1758-2946-3-33>

Pacini, C., Iorio, F., Gonçalves, E., Iskar, M., Klabunde, T., Bork, P., & Saez-Rodriguez, J. (2013). DvD: An R/Cytoscape pipeline for drug repurposing using public repositories of gene expression data. *Bioinformatics*, 29(1), 132–134. <https://doi.org/10.1093/bioinformatics/bts656>

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In S. Markidis & E. Laure (Eds.), *Solving Software Challenges for Exascale* (Vol. 8759, pp. 3–27). Springer International Publishing. https://doi.org/10.1007/978-3-319-15976-8_1

Paul, S. M., & Lewis-Hall, F. (2013). Drugs in Search of Diseases. *Science Translational Medicine*, 5(186). <https://doi.org/10.1126/scitranslmed.3004452>

Peng, J., & Xu, J. (2009). Boosting Protein Threading Accuracy. In S. Batzoglou (Ed.), *Research in Computational Molecular Biology* (Vol. 5541, pp. 31–45). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-02008-7_3

Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng, E. C., & Ferrin, T. E. (2004). UCSF Chimera—A visualization system for exploratory research and analysis. *Journal of Computational Chemistry*, 25(13), 1605–1612. <https://doi.org/10.1002/jcc.20084>

Pogliani, L. (2000). From Molecular Connectivity Indices to Semiempirical Connectivity Terms: Recent Trends in Graph Theoretical Descriptors. *Chemical Reviews*, 100(10), 3827–3858. <https://doi.org/10.1021/cr0004456>

Polamreddy, P., & Gattu, N. (2019). The drug repurposing landscape from 2012 to 2017: Evolution, challenges, and possible solutions. *Drug Discovery Today*, 24(3), 789–795. <https://doi.org/10.1016/j.drudis.2018.11.022>

Porta-Pardo, E., Ruiz-Serra, V., Valentini, S., & Valencia, A. (2022). The structural coverage of the human proteome before and after AlphaFold. *PLOS Computational Biology*, 18(1), e1009818. <https://doi.org/10.1371/journal.pcbi.1009818>

Porter, K. A., Desta, I., Kozakov, D., & Vajda, S. (2019). What method to use for protein–protein docking? *Current Opinion in Structural Biology*, 55, 1–7. <https://doi.org/10.1016/j.sbi.2018.12.010>

Prada Gori, D. N., Alberca, L. N., & Talevi, A. (2023). Making the most effective use of available computational methods for drug repositioning. *Expert Opinion on Drug Discovery*, 18(5), 495–503. <https://doi.org/10.1080/17460441.2023.2198700>

Priotto, G., Kasparian, S., Mutombo, W., Ngouama, D., Ghorashian, S., Arnold, U., Ghabri, S., Baudin, E., Buard, V., Kazadi-Kyanza, S., Ilunga, M., Mutangala, W., Pohlig, G., Schmid, C., Karunakara, U., Torreele, E., & Kande, V. (2009). Nifurtimox-eflornithine combination therapy for second-stage African Trypanosoma brucei gambiense trypanosomiasis: A multicentre, randomised, phase III, non-inferiority trial. *The Lancet*, 374(9683), 56–64. [https://doi.org/10.1016/S0140-6736\(09\)61117-X](https://doi.org/10.1016/S0140-6736(09)61117-X)

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>

Rational Discovery LLC. (2006). *RDKit: Open-Source Cheminformatics and Machine Learning Software, Open-Source Cheminformatics and Machine Learning*. <http://www.rdkit.org/>

Rost, B. (1999). Twilight zone of protein sequence alignments. *Protein Engineering, Design and Selection*, 12(2), 85–94. <https://doi.org/10.1093/protein/12.2.85>

Rudrapal, M., Khairnar, S. J., & Jadhav, A. G. (2020). Drug repurposing (DR): an emerging approach in drug discovery. *Drug repurposing-hypothesis, molecular aspects and therapeutic applications*, 10.

Sbaraglini, M., Cristina Vanrell, M., Leticia Bellera, C., Benaim, G., Carrillo, C., Talevi, A., & Silvia Romano, P. (2016). Neglected Tropical Protozoan Diseases: Drug

Repositioning as a Rational Option. *Current Topics in Medicinal Chemistry*, 16(19), 2201–2222. <https://doi.org/10.2174/1568026616666160216154309>

Schneidman-Duhovny, D., Dror, O., Inbar, Y., Nussinov, R., & Wolfson, H. J. (2008). PharmaGist: A webserver for ligand-based pharmacophore detection. *Nucleic Acids Research*, 36(Web Server), W223–W228. <https://doi.org/10.1093/nar/gkn187>

Service, R. F. (2020). ‘The game has changed.’ AI triumphs at protein folding. *Science*, 370(6521), 1144–1145. <https://doi.org/10.1126/science.370.6521.1144>

Singhal, S., Mehta, J., Desikan, R., Ayers, D., Roberson, P., Eddlemon, P., Munshi, N., Anaissie, E., Wilson, C., Dhodapkar, M., Zeldis, J., Siegel, D., Crowley, J., & Barlogie, B. (1999). Antitumor Activity of Thalidomide in Refractory Multiple Myeloma. *New England Journal of Medicine*, 341(21), 1565–1571. <https://doi.org/10.1056/NEJM199911183412102>

Small, J. C., Joblin-Mills, A., Carbone, K., Kost-Alimova, M., Ayukawa, K., Khodier, C., Dancik, V., Clemons, P. A., Munkacsi, A. B., & Wagner, B. K. (2022). Phenotypic Screening for Small Molecules that Protect β -Cells from Glucolipototoxicity. *ACS Chemical Biology*, 17(5), 1131–1142. <https://doi.org/10.1021/acscchembio.2c00052>

Sotriffer, C. (Ed.). (2011). *Virtual screening: Principles, challenges, and practical guidelines*. Wiley-VCH.

Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>

Swain, M. (2019). *MolVS: Molecule Validation and Standardization—MolVS 0.1.1 documentation*. <https://molvs.readthedocs.io/en/latest/>

Swinney, D. C. (2013). Phenotypic vs. Target-Based Drug Discovery for First-in-Class Medicines. *Clinical Pharmacology & Therapeutics*, 93(4), 299–301. <https://doi.org/10.1038/clpt.2012.236>

Talevi, A. (2011). *Virtual Screening: An Emergent, Key Methodology for Drug Development in an Emergent Continent-A Bridge Towards Patentability* (pp. 229–245). In *Advanced Methods and Applications in Chemoinformatics: Research Progress and New Applications*. <https://www.semanticscholar.org/paper/Virtual-Screening-%3A-An-Emergent-%2C-Key-Methodology-Talevi/b1f67cc3b4b41c0ecdbdfce684a7e7aea31da40e>

Talevi, A. (2023). Computer-Aided Drug Discovery and Design: Recent Advances and Future Prospects. In M. Gore & U. B. Jagtap (Eds.), *Computational Drug*

Discovery and Design (Vol. 2714, pp. 1–20). Springer US.
https://doi.org/10.1007/978-1-0716-3441-7_1

Talevi, A., Carrillo, C., & Comini, M. (2018). The Thiol-polyamine Metabolism of *Trypanosoma cruzi*: Molecular Targets and Drug Repurposing Strategies. *Current Medicinal Chemistry*, 26(36), 6614–6635.
<https://doi.org/10.2174/0929867325666180926151059>

Talevi, Alan & Bruno-Blanch, Luis E. (2009). Screening Virtual: Una Herramienta Eficaz para el Desarrollo de Nuevos Fármacos en Latinoamérica. *Latin American Journal of Pharmacy*.

Temesi, G., Bolgár, B., Arany, Á., Szalai, C., Antal, P., & Mátyus, P. (2014). Early repositioning through compound set enrichment analysis: A knowledge-recycling strategy. *Future Medicinal Chemistry*, 6(5), 563–575.
<https://doi.org/10.4155/fmc.14.4>

Teo, S. K., Resztak, K. E., Scheffler, M. A., Kook, K. A., Zeldis, J. B., Stirling, D. I., & Thomas, S. D. (2002). Thalidomide in the treatment of leprosy. *Microbes and Infection*, 4(11), 1193–1202. [https://doi.org/10.1016/S1286-4579\(02\)01645-3](https://doi.org/10.1016/S1286-4579(02)01645-3)

Terry, C. & Lesser, N. (2015). *Measuring the return from pharmaceutical innovation 2015*. Deloitte Turkey. Retrieved September 14, 2023, from <https://www2.deloitte.com/tr/en/pages/life-sciences-and-healthcare/articles/measuring-return-from-pharmaceutical-innovation1.html>

The AlphaFold team. (2022). *AlphaFold*. Retrieved September 14, 2023, from <https://www.deepmind.com/research/highlighted-research/alphafold>

The AlphaFold team. (2020). *AlphaFold: A solution to a 50-year-old grand challenge in biology*. Retrieved September 14, 2023, from <https://www.deepmind.com/blog/alphafold-a-solution-to-a-50-year-old-grand-challenge-in-biology>

The AlphaFold Team & Google Colab. (2022). *ColabFold*. <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=G4yBrceuFbf3>

The Investopedia Team. (2022). *Average Research and Development Costs for Pharmaceutical Companies*. Investopedia. Retrieved October 9, 2023, from <https://www.investopedia.com/ask/answers/060115/how-much-drug-companys-spending-allocated-research-and-development-average.asp>

Todeschini, R., & Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics* (1st ed.). Wiley. <https://doi.org/10.1002/9783527628766>

Torreele, E., Bourdin Trunz, B., Tweats, D., Kaiser, M., Brun, R., Mazué, G., Bray, M. A., & Pécou, B. (2010). Fexinidazole – A New Oral Nitroimidazole Drug Candidate Entering Clinical Development for the Treatment of Sleeping Sickness. *PLoS Neglected Tropical Diseases*, 4(12), e923. <https://doi.org/10.1371/journal.pntd.0000923>

Tuckerman, M. E., & Martyna, G. J. (2000). Understanding Modern Molecular Dynamics: Techniques and Applications. *The Journal of Physical Chemistry B*, 104(2), 159–178. <https://doi.org/10.1021/jp992433y>

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701–1718. <https://doi.org/10.1002/jcc.20291>

Vidović, Koletić, A., & Schreiber, S. C. (2014). Large-scale integration of small molecule-induced genome-wide transcriptional responses, Kinome-wide binding affinities and cell-growth inhibition profiles reveal global trends characterizing systems-level drug action. *Frontiers in Genetics*, 5. <https://doi.org/10.3389/fgene.2014.00342>

Walters, W. P., Stahl, M. T., & Murcko, M. A. (1998). Virtual screening—An overview. *Drug Discovery Today*, 3(4), 160–178. [https://doi.org/10.1016/S1359-6446\(97\)01163-X](https://doi.org/10.1016/S1359-6446(97)01163-X)

Weng, H.-B., Chen, H.-X., & Wang, M.-W. (2018). Innovation in neglected tropical disease drug discovery and development. *Infectious Diseases of Poverty*, 7(1), 67. <https://doi.org/10.1186/s40249-018-0444-1>

Wermuth, C. G., Ganellin, C. R., Lindberg, P., & Mitscher, L. A. (1998). Glossary of terms used in medicinal chemistry (IUPAC Recommendations 1998). *Pure and Applied Chemistry*, 70(5), 1129–1143. <https://doi.org/10.1351/pac199870051129>

Willett, P. (2006). Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, 11(23–24), 1046–1053. <https://doi.org/10.1016/j.drudis.2006.10.005>

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>

Yang, S.-Y. (2010). Pharmacophore modeling and applications in drug discovery: Challenges and recent advances. *Drug Discovery Today*, 15(11–12), 444–450. <https://doi.org/10.1016/j.drudis.2010.03.013>

Yao, H., Liu, J., Xu, M., Ji, J., Dai, Q., & You, Z. (2022). Discussion on molecular dynamics (MD) simulations of the asphalt materials. *Advances in Colloid and Interface Science*, 299, 102565. <https://doi.org/10.1016/j.cis.2021.102565>

Zhou, S., Wang, F., Hsieh, T.-C., Wu, J., & Wu, E. (2013). Thalidomide—A Notorious Sedative to a Wonder Anticancer Drug. *Current Medicinal Chemistry*, 20(33), 4102–4108. <https://doi.org/10.2174/09298673113209990198>

CAPÍTULO 3: METODOLOGÍA

3.1 INTRODUCCIÓN

El presente capítulo se encuentra dividido en tres bloques temáticos: las secciones 3.2 a 3.7 inclusive se refieren a la metodología implementada en la generación y validación de modelos **QSAR-2D** (métodos indirectos), mientras que la sección 3.8 aborda tanto al modelado estructural de las **FABPs** de cestodos como al cribado virtual por *docking* molecular (métodos directos). Finalmente, la sección 3.9 describe la metodología experimental utilizada para la evaluación *in vitro* de los candidatos obtenidos *in silico* durante ambas campañas de **CV**.

3.2 MÉTODOS INDIRECTOS. GENERACIÓN DEL MODELO QSAR

La actividad biológica A de un compuesto se puede expresar como una función de características estructurales codificadas en un número d de descriptores moleculares (Grisoni et al., 2018). Consecuentemente, la actividad A de un compuesto no testeado puede ser inferida a partir de su estructura molecular, mediante el cálculo de los descriptores moleculares seleccionados (X_i), tal como se muestra en la **Ecuación 3.1**.

$$A = f(X_1, X_2, X_3, \dots, X_d) \quad (3.1)$$

Esta metodología se conoce como **QSAR**, ya presentada en el capítulo anterior (Davis, 2014). Según la naturaleza de la variable dependiente considerada, se pueden construir modelos **QSAR** cuantitativos y cualitativos (funciones discriminantes).

El proceso de desarrollo de modelos **QSAR** es complejo e implica varios pasos, incluyendo la compilación y curado de una base de datos de compuestos que cuya actividad A ha sido medida experimentalmente, el cálculo de los descriptores moleculares, la extracción de información y generación de modelos y la validación de los mismos.

Uno de los objetivos específicos de este trabajo de tesis involucra el desarrollo de funciones discriminantes capaces de identificar inhibidores del transporte e incorporación de lípidos en los parásitos cestodos *E. granulosus*, *E. multilocularis* y *T. solium* en bases de datos de compuestos químicos, debido a la ya mencionada

capacidad limitada de estos parásitos de sintetizar compuestos hidrofóbicos *de novo*.

A continuación, la **Figura 3.2** muestra un esquema del flujo de trabajo utilizado en la generación, validación y aplicación de los modelos en la campaña de **CV** basada desde el ligando.

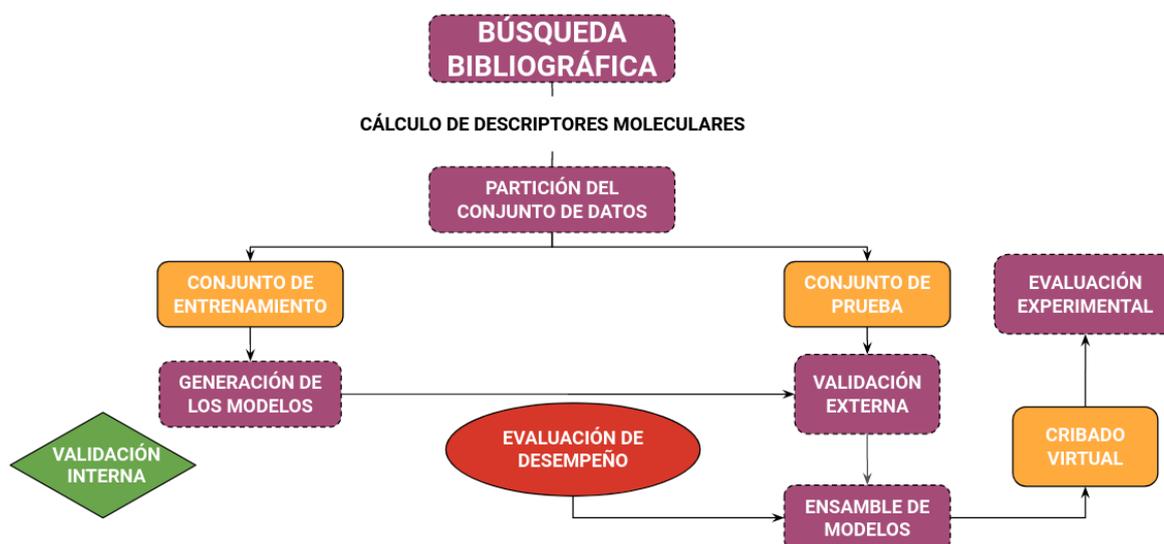


Figura 3.2. Flujo de trabajo del desarrollo de los modelos **QSAR-2D**.

3.2.1 COMPILADO DEL CONJUNTO DE DATOS

La generación de modelos **QSAR** se inicia con la obtención o recopilación de datos sobre la actividad de compuestos químicos sobre el blanco molecular de interés. En los inicios de este proyecto de tesis doctoral se sabía de la escasez de compuestos con actividad reportada contra las **FABPs** de cestodos, con lo cual se decidió recurrir al enfoque de reposicionamiento de blancos (Klug et al., 2016), ya descrito en el capítulo 2. A su vez, debido al éxito en la campaña de **CV** reportada por Bélgamo et al. (2020), en donde se evidencia la unión de tres compuestos de reposicionamiento a tres isoformas de **FABP** de *E. multilocularis* a partir de modelos **QSAR** generados con moléculas inhibitoras de **AFABP**, se decidió utilizar la misma estrategia, expandiendo el conjunto de datos previamente compilado por el grupo de trabajo. Es menester recordar que la **AFABP** es la isoforma que cuenta con mayor similitud estructural con respecto a la de cestodos, como ya se desarrolló en el capítulo 1.

La compilación del conjunto de datos es un paso crítico y fundamental para obtener modelos **QSAR** de alta calidad. Para satisfacer este requisito de manera óptima, deben tenerse en cuenta los siguientes aspectos (Davis, 2014; Tropsha, 2010; Jaworska et al., 2005):

1. Los compuestos deben haber sido evaluados en un ensayo *in vitro*, determinando la capacidad de unión al blanco molecular. Idealmente, se preferirá que los compuestos hayan sido evaluados bajo las mismas condiciones experimentales.
2. El método de determinación de la actividad debe tener poco error asociado.
3. Contar con una gran diversidad estructural dentro del conjunto de datos, para que los modelos puedan ser aplicados a espacios químicos más amplios.
4. Los valores de actividad reportada deberían presentar una distribución adecuada entre tres y cuatro órdenes logarítmicos.

Como ya se mencionó, el conjunto de datos en nuestro caso se compiló a partir de búsqueda de compuestos con actividad reportada contra **AFABP**. Para ello, se utilizaron los repositorios Scopus (Elsevier, 2023) y PubMed (*National Institutes of Health*, 2023). En ambos casos, las palabras clave fueron "**FABP**" e "*inhibitor*", utilizando el conector booleano *AND*. La búsqueda se limitó hasta noviembre del año 2021.

Debido a que los datos de actividad fueron obtenidos de diferentes trabajos científicos, las condiciones experimentales no son idénticas, con lo cual se utilizó como técnica de modelado al Análisis Lineal Discriminante, debido a que un modelo categórico o clasificador permite amortiguar el efecto de la heterogeneidad en la obtención experimental de los datos (Talevi et al., 2011).

Los compuestos compilados fueron clasificados como **ACTIVOS** o **INACTIVOS** en función de su **IC₅₀** ó **Ki** contra **AFABP**:

- **ACTIVOS**: $IC_{50} / Ki \leq 10 \mu M$
- **INACTIVOS**: $IC_{50} / Ki > 20 \mu M$

De esta manera, compuestos cuya **IC₅₀** ó **Ki** se encontraba en el rango de 10-20 μM no fueron tenidos en cuenta para ninguna de las categorías, a fines de reducir la posibilidad de asignar erróneamente moléculas a la clase incorrecta.

El conjunto de datos compilado de literatura fue de 288 compuestos químicos, siendo 187 considerados **ACTIVOS** mientras que 101 fueron considerados **INACTIVOS**.

3.2.2 PARTICIÓN DEL CONJUNTO DE DATOS

En la construcción de modelos **QSAR**, el conjunto de datos se dividirá típicamente en al menos dos subconjuntos: el conjunto de entrenamiento, utilizado para generar los modelos, y el conjunto de prueba, necesario para validarlos. Esta división tiene la ventaja de que los compuestos del conjunto de prueba son desconocidos por los modelos, ya que han sido excluidos del proceso de entrenamiento de los mismos (Martin et al., 2012). En el contexto de la presente tesis doctoral, el conjunto de

entrenamiento está formado por 202 compuestos (101 ACTIVOS y 101 INACTIVOS), mientras que los 86 ACTIVOS restantes fueron utilizados para la generación de compuestos presuntamente INACTIVOS, los cuales se conocen como moléculas señuelo o *decoys*. Estas moléculas señuelo, junto con los 86 ACTIVOS recién mencionados forman parte de las quimiotecas simuladas que hacen las veces de conjunto de prueba. La generación de señuelos se detalla en la sección 3.3.4.

Hawkins et al. (2008) postulan que mayor cantidad de compuestos disponibles para el entrenamiento y validación permitirán obtener modelos más robustos. Si el conjunto de entrenamiento es demasiado pequeño, hay mayor probabilidad de observar correlaciones espurias y sobreajuste, lo que perjudicará la capacidad predictiva de los modelos. En el mismo sentido, Tropsha (2010) considera que una cantidad adecuada de compuestos en el conjunto total de datos se encuentra entre los 150-300 compuestos, en donde el conjunto de entrenamiento debe contar con la misma cantidad de cada clase o categoría. Por estos motivos, se decidió contar con un conjunto de entrenamiento balanceado incluyendo a todos los compuestos INACTIVOS reales, mientras que todos los compuestos ACTIVOS fueron particionados de manera racional.

Existen varias estrategias para la partición de los datos. Una de ellas es la asignación al azar de los compuestos del set de datos a cada uno de los conjuntos. Esta estrategia es adecuada para sets de datos de tamaño tal que permita seleccionar conjuntos de prueba y de entrenamiento de tamaño similar (Golbraikh et al., 2003; Golbraikh & Tropsha, 2000). La principal desventaja de este método es que, al no proporcionar un criterio racional para la asignación de las moléculas a cada conjunto de datos, no se puede garantizar que la partición del conjunto de datos sea representativa, es decir que el espacio químico cubierto por los compuestos de los conjuntos de entrenamiento y de prueba sea similar. La partición no representativa del conjunto de datos podría ocasionar que los compuestos del conjunto de prueba no sean adecuados para validar los modelos generados con el conjunto de entrenamiento (ya sea porque contienen casos que caen fuera del dominio de aplicación o porque no consiguen abarcar toda la diversidad química del conjunto de entrenamiento). La probabilidad de que surja este problema aumenta cuanto menor es el tamaño del conjunto de datos (Hawkins et al., 2008).

Para enmendar esta situación se han desarrollado algoritmos de partición racional que intentan dividir los datos representativamente, de manera de poder estimar con mayor confianza la capacidad predictiva de los modelos generados. En la siguiente sección se describe la estrategia de partición representativa utilizada en este trabajo.

3.2.2.1 ESTRATEGIA DE PARTICIÓN UTILIZADA EN ESTE TRABAJO

Para garantizar una partición representativa de los compuestos **ACTIVOS** para el entrenamiento y las quimiotecas simuladas, se aplicó una metodología agrupamiento (*clustering* en inglés) de tipo no jerárquica.

Los métodos de *clustering* son técnicas de clasificación no supervisada, lo que significa que no se requiere una variable dependiente asignada por el usuario para su aplicación. Estos métodos permiten agrupar los datos solo en base a las características moleculares de los compuestos. En el campo del descubrimiento de fármacos, los métodos de *clustering* se han utilizado ampliamente para clasificar las moléculas según su quimiotipo. Existen métodos basados en huellas dactilares moleculares (en inglés *molecular fingerprints*) y en descriptores moleculares, que se clasifican comúnmente en dos clases: los métodos jerárquicos, como el agrupamiento de Ward (Ward, 1963), y los métodos no jerárquicos, como el de Jarvis-Patrick (Jarvis & Patrick, 1973) y *k-means* (en español, k-medias) (Böcker, 2008). En este trabajo de tesis, se utilizó **iRaPCA** (del inglés *iterative Random subspace Principal Component Analysis clustering*), un algoritmo iterativo basado en la combinación de subespacios aleatorios y el agrupamiento por *k-means*. El método contempla el cálculo de descriptores moleculares del conjunto de datos y su posterior reducción dimensional mediante análisis de componentes principales (**PCA**, por sus siglas en inglés) (Prada Gori et al., 2022a).

El algoritmo **iRaPCA** se puede describir en tres pasos:

1. **Entrada y codificación:** El conjunto de datos de moléculas a agrupar se proporciona en un archivo de extensión .csv, en notación **SMILES** (acrónimo del anglicismo *simplified molecular-input line-entry system*) (Weininger, 1988). Cada una de las moléculas en formato **SMILES** es estandarizada, asignando los isótopos más comunes para el fragmento más grande de la molécula, eliminando cargas y estereoquímica. Luego, se calculan un total de 1613 descriptores independientes de la conformación (0D-2D) utilizando el paquete Mordred (Moriwaki et al., 2018).
2. **Espacio de descriptores y reducción de dimensionalidad:** Del conjunto de 1613 descriptores independientes de la conformación, se eliminan aquellos cuya varianza sea menor al valor umbral seleccionado, ya que los descriptores con baja varianza no proporcionan información relevante. Luego, se exploran subespacios aleatorios (una descripción detallada se provee en la sección 3.2.4) para explorar el conjunto de descriptores remanentes. Seguidamente, los descriptores altamente correlacionados se eliminan de cada subconjunto de descriptores aleatorios, evitando así tener descriptores con información redundante. Como las distancias euclídeas serán medidas durante el procedimiento de *clustering*, y este es sensible a la escala

numérica de cada uno de los descriptores, se realiza una normalización de los descriptores previa al agrupamiento. Para ello, se utiliza la función `MinMaxScaler` del módulo de Python `scikit-learn` (Pedregosa et al., 2011). Finalmente, se realiza la reducción de la dimensionalidad vía **PCA** para cada subconjunto para obtener las dos primeras componentes principales.

3. **Clustering:** Se aplica el algoritmo *k-means* a cada subconjunto, reducido a las primeras dos componentes principales, variando sistemáticamente el número de clusters (**k**) desde 2 hasta un valor seleccionado por el usuario. La implementación del algoritmo en `scikit-learn` selecciona 10 centroides pseudoaleatorios diferentes por *cluster*, proporcionando la mejor salida en términos de inercia (o, lo que es equivalente, suma de cuadrados intra-cluster). Por defecto, el valor máximo de **k** se establece en 25. Los *clusters* recién formados se evalúan calculando el valor medio del coeficiente *Silhouette* (**SIL**) (Rousseeuw, 1987) para cada valor de **k** y cada par de subconjunto de descriptores. Los resultados se presentan en un gráfico **SIL** vs **k**. Se selecciona el subconjunto y el valor de **k** que proporcionan el puntaje **SIL** más alto para agrupar a las moléculas.

Tras la primera ronda de *clustering*, se calcula para cada cluster la relación entre el número de moléculas asignadas al *cluster* y el total de moléculas en el conjunto de datos. Los grupos que superan un valor de corte de esta relación (definido por el usuario) se someten a una nueva ronda de agrupamiento con los mismos parámetros iniciales. Los grupos que no superan el valor de corte seleccionado se mantienen tal como están. Por defecto, esta proporción se establece en 0,3. Cuando ya no hay más grupos que superen el límite de proporción o cuando el algoritmo alcanza el número máximo de iteraciones fijado por el usuario, el procedimiento de *clustering* finaliza, obteniendo como resultado la lista de moléculas asignadas a cada *cluster* y *subcluster*.

Los parámetros utilizados para el agrupamiento del conjunto de datos fueron:

- Valor de corte para la varianza de los descriptores: 0,05
- Número de subespacios: 100
- Número de descriptores por subespacio: 200
- Coeficiente de correlación: Pearson
- Punto de corte para el coeficiente de correlación: 0,4
- Número mínimo de descriptores por subespacio: 4
- Número máximo de descriptores por subespacio: 25
- Número mínimo de *clusters* por iteración: 2
- Número máximo de *clusters* por iteración: 25
- Máxima relación moléculas *clusters*/total de moléculas: 0,3
- Número de iteraciones: 1
- Número de componentes principales a utilizar: 2

3.2.3 CÁLCULO DE DESCRIPTORES MOLECULARES

Como ya se mencionó, los descriptores moleculares dependientes de la conformación requieren de la pre-optimización de la geometría molecular, lo cual es muy demandante computacionalmente para la generación de modelos **QSAR**. Por este motivo, se generaron modelos **QSAR** utilizando descriptores de baja dimensionalidad, es decir, aquellos que no dependen de la conformación de las moléculas (0D-2D).

Las moléculas del conjunto de datos compilado fueron dibujadas en la interfaz gráfica de ChEMBL (Davies et al., 2015; Gaulton et al., 2017), que utiliza el programa MarvinJS (Chemaxon, 2023) como generador de estructuras químicas. Para cada compuesto se obtuvo la notación **SMILES** (Weininger, 1988) presente en ChEMBL y se estandarizaron utilizando la herramienta Standarizer presente en la biblioteca MolVS de Python (Swain, 2019). Para el caso de los señuelos, los mismos son obtenidos en formato **SMILES** directamente, y estandarizados de manera análoga a los ACTIVOS e INACTIVOS. Muy brevemente, en la estandarización de los códigos **SMILES** se busca uniformar el formato de todos los compuestos presentes en el conjunto de datos. La estandarización realizada consistió en:

- Tomar el fragmento covalente más grande de la molécula (eliminando así las sales).
- Eliminar la estereoisomería de los compuestos.
- Eliminar la tautomería, obteniendo el tautómero canónico.
- Eliminar la carga de la molécula, reemplazando por la versión sin carga de la misma.
- Reemplazar todos los átomos por el isótopo más abundante para cada elemento.

Una vez estandarizado el conjunto de datos, se calcularon 1613 descriptores independientes de la conformación (0D-2D) utilizando el paquete Mordred (Moriwaki et al., 2018).

Los descriptores que resultaron ser constantes y los que presentaban baja variabilidad (desviación estándar < 0,05) para los compuestos del conjunto de entrenamiento fueron excluidos.

3.2.4 MODELADO

Como ya se mencionó, los modelos clasificatorios son útiles para el tratamiento de datos ruidosos, permitiendo en buena medida utilizar datos experimentales colectados mediante distintos protocolos, condiciones experimentales y/o experimentadores (Talevi et al., 2011). Al utilizar una variable dependiente observada binaria o discreta, el modelo se enfocará en la tarea de clasificación, mitigándose así el ruido que podría provenir del error o variabilidad hacia dentro de

cada categoría de ejemplos de entrenamiento (excepto que tal error o variabilidad pudiera conducir a una etiqueta errónea de los datos).

En nuestro caso, los compuestos del conjunto de datos fueron clasificados de manera categórica en dos clases: **ACTIVOS** e **INACTIVOS**, asignando un valor arbitrario de 1 y 0, respectivamente, para cada una de ellas. De esta manera, esta etiqueta binaria es adoptada como el valor observado de la variable dependiente en la generación de modelos. Cabe recordar que los compuestos señuelo son considerados compuestos **INACTIVOS**.

3.2.4.1 SUBESPACIOS ALEATORIOS

Los descriptores moleculares fueron divididos aleatoriamente en 4000 subconjuntos con 200 descriptores cada uno, generando un modelo independiente a partir de cada subconjunto.

La técnica de subespacios aleatorios consiste en la generación de modelos en paralelo, en donde los descriptores moleculares, que hacen las veces de variables independientes de los modelos, son muestreados al azar y cada modelo se entrena en un subconjunto de descriptores en particular (Daho & Chikh, 2015; Vyskovsky et al., 2016). Esto conlleva que cada modelo no se enfocará necesariamente en descriptores con alta capacidad explicativa en el conjunto de entrenamiento, ganando así capacidad predictiva y reduciendo el sobreajuste.

3.2.4.2 SELECCIÓN DE DESCRIPTORES Y GENERACIÓN DE MODELOS

Para llevar a cabo la selección de los descriptores que mejor correlacionan con la variable dependiente en cada uno de los subconjuntos se utilizó la técnica de semi correlación por pasos hacia adelante (en inglés *Forward Stepwise*) (Toropova & Toropov, 2017), implementada en Python.

Esta técnica es un método habitual de selección de variables en el análisis de regresión. El procedimiento comienza con un modelo sin variables, sumando de a una variable por vez, de acuerdo a la capacidad de cada variable para mejorar la capacidad explicativa del modelo. En cada paso, se evalúa cómo mejora el ajuste del modelo con cada una de las variables remanentes en el pool de descriptores, seleccionando aquella que produce una mayor ganancia en términos explicativos. Este proceso se repite hasta que se han agregado todas las variables predictoras al modelo o hasta que el modelo alcanza un nivel satisfactorio de ajuste (Choueiry, 2022). El modelo final obtenido de esta manera depende críticamente de la variable independiente añadida en el primer paso del procedimiento, que condiciona fatídicamente todas las demás, conduciendo a resultados previsiblemente subóptimos. Sin embargo, esta desventaja es superada cuando la técnica se

combina con una exploración aleatoria de subconjuntos del pool de descriptores, variándose así la variable que se incorpora al modelo en el primer paso del procedimiento.

Se configuró un nivel de significancia de 0,05 ($pvalue=0,05$), para que una variable pueda ser incorporada al modelo, mientras que se fijaron 5 descriptores a incorporar como máximo a cada modelo de regresión, a fin de evitar el sobreajuste.

El término sobreajuste se refiere al entrenamiento excesivo de un algoritmo, es decir dotar al modelo de un alto poder explicativo sobre el conjunto de entrenamiento (en analogía al aprendizaje humano, podría pensarse como una memorización de contenidos) a expensas de perder capacidad de predicción sobre el conjunto de prueba o sobre otros datos externos. Se recomienda una relación de al menos 15 entre número de instancias de entrenamiento y la cantidad de descriptores incluidos en el modelo, o en otras palabras, no debe incorporarse más de un predictor por cada 15 moléculas presentes en el conjunto de entrenamiento (Peduzzi et al., 1996; Peduzzi et al., 1995).

Los modelos o funciones discriminantes obtenidas (FD) se corresponden a planos o hiperplanos k -dimensionales (siendo k el número de descriptores por modelo) que dejan ejemplos de cada categoría (ACTIVOS e INACTIVOS) a uno u otro lado del plano o hiperplano. La **Ecuación 3.2** representa de manera genérica a una FD lineal, donde el valor de a_0 representa a la ordenada al origen, mientras que a_i se corresponde con el coeficiente i para el descriptor d_i del modelo generado.

$$FD = a_0 + \sum_{i=1}^k a_i * d_i \quad (3.2)$$

3.3 VALIDACIÓN DE LOS MODELOS QSAR

En el desarrollo de modelos **QSAR**, la validación es una parte central del procedimiento, que evalúa la capacidad predictiva y robustez de los modelos (Golbraikh et al., 2003). Los métodos de validación se dividen en dos categorías: validación interna, que utiliza los compuestos del propio conjunto de entrenamiento para estimar la capacidad predictiva y la robustez, y validación externa, que utiliza para esos mismos fines compuestos no utilizados en el entrenamiento de los modelos. Tropsha (2010) ha destacado la importancia de utilizar ambos métodos para seleccionar modelos de alta calidad. Sin embargo, Hawkins et al. (2003) sugieren que la validación externa puede ser omitida en casos donde el número de compuestos es limitado.

3.3.1 MÉTODOS DE VALIDACIÓN INTERNA

Se generaron modelos y se validaron mediante dos métodos de validación interna: la validación cruzada *Leave Group Out* (**LGO**) y aleatorización de Fisher, también conocido como *y-Randomization* en inglés.

El método **LGO** se utilizó para determinar la robustez de los modelos al identificar si algunos de los compuestos del conjunto de entrenamiento influyen excesivamente en los parámetros del modelo inferido. Este método consiste en remover un número n de compuestos del conjunto de entrenamiento, generar un modelo con los compuestos restantes y evaluar el nuevo modelo en el conjunto de compuestos removido. Este proceso se repite hasta que todos los compuestos hayan sido removidos y evaluados al menos una vez (Kunal & Supratik, 2015). Para la implementación de **LGO** en este trabajo, se removieron 20 compuestos del conjunto de entrenamiento por ronda (10 ACTIVOS y 10 INACTIVOS, 1000 rondas). Los resultados de **LGO** se informan como el promedio de los porcentajes de buenas clasificaciones en 1000 rondas de validación, junto con su desvío estándar.

Por otro lado, la aleatorización de Fisher se empleó para descartar la posibilidad de una correlación aleatoria entre las variables independientes (descriptores) y la variable dependiente (actividad). Este método busca eliminar cualquier relación entre las estructuras de los compuestos del conjunto de entrenamiento y su actividad, a fin de determinar si el proceso de modelado utilizado produce modelos espurios con parámetros estadísticos aceptables (Rücker et al., 2007).

Para llevar a cabo la aleatorización de Fisher, se aleatorizaron los valores observados de la variable dependiente de los compuestos del conjunto de entrenamiento. En consecuencia, algunos de los compuestos activos se etiquetaron como inactivos y viceversa. Luego se generan nuevamente los modelos. Se espera que los modelos generados de esta manera tengan un desempeño estadístico deficiente y una capacidad de predicción limitada. Los resultados de esta validación se expresan como el promedio de los porcentajes de buenas clasificaciones obtenidos al aleatorizar 1000 veces la variable dependiente y su correspondiente desviación estándar.

3.3.2 VALIDACIÓN EXTERNA

La validación externa es un proceso en el que se evalúa la capacidad predictiva de los modelos **QSAR** en un conjunto de moléculas que no se utilizaron para su generación. Aunque los métodos de validación interna pueden arrojar buenos resultados, se ha observado que los mismos tienden a ser sobre-optimistas (Veerasingam et al., 2011). La capacidad predictiva de los modelos **QSAR** solo puede establecerse fehacientemente a través de la aplicación exitosa de los modelos para predecir la actividad sobre un conjunto de datos externos de tamaño adecuado.

Para llevar a cabo esta validación, se utilizaron los modelos generados a partir del conjunto de entrenamiento para predecir la actividad de los compuestos en las quimiotecas simuladas.

La capacidad predictiva de los modelos individuales se mide mediante el valor del área bajo la curva Característica Operativa del Receptor (en inglés *area under the Receiver Operating Characteristics curve*, **AUCROC**), estimador que es abordado en la siguiente sección.

3.3.3 EVALUACIÓN DE LOS MODELOS

Para evaluar el desempeño de un modelo **QSAR**, se utilizan dos indicadores clave: la sensibilidad (**Se**), que mide la tasa de verdaderos positivos, y la especificidad (**Sp**), que mide la tasa de verdaderos negativos. Estos valores se calculan utilizando las **Ecuaciones 3.3**, donde **FN** representa falsos negativos, **FP** representa falsos positivos, **VN** representa verdaderos negativos y **VP** representa verdaderos positivos.

$$Se = \frac{VP}{VP+FN} \quad (I) \quad Sp = \frac{VN}{VN + FP} \quad (II) \quad (3.3)$$

Después de aplicar el modelo a cada molécula del conjunto de datos, se obtiene un puntaje numérico para cada una de ellas. Para determinar si cada compuesto es ACTIVO o INACTIVO, se establece un umbral de puntuación. Sin embargo, al modificar este umbral, la **Se** y la **Sp** varían inversamente, por lo que es necesario encontrar un equilibrio adecuado entre ambos. Las ya mencionadas curvas **ROC** (Característica Operativa del Receptor, en inglés *Receiver Operating Characteristics*) se utilizan para encontrar este equilibrio. Las curvas **ROC** son gráficos que muestran la relación entre la **Se** y $1 - Sp$, y permiten determinar el umbral óptimo para clasificar los compuestos (**Figura 3.3.3a**).

El **AUCROC** es una medida valiosa para evaluar el desempeño de un modelo **QSAR** en comparación con una clasificación aleatoria (Triballeau et al., 2005). Un **AUCROC** de 0,5 corresponde a una clasificación aleatoria mientras que un **AUCROC** de 1 corresponde a un modelo perfecto. La capacidad de predicción del modelo aumenta a medida que el **AUCROC** se aleja de la línea diagonal (**Figura 3.3.3d**). Para evaluar el desempeño de los modelos **QSAR**, se calcularon los **AUCROC** utilizando el módulo `roc_auc_score` de la biblioteca `scikit-learn` de Python (Pedregosa et al., 2011).

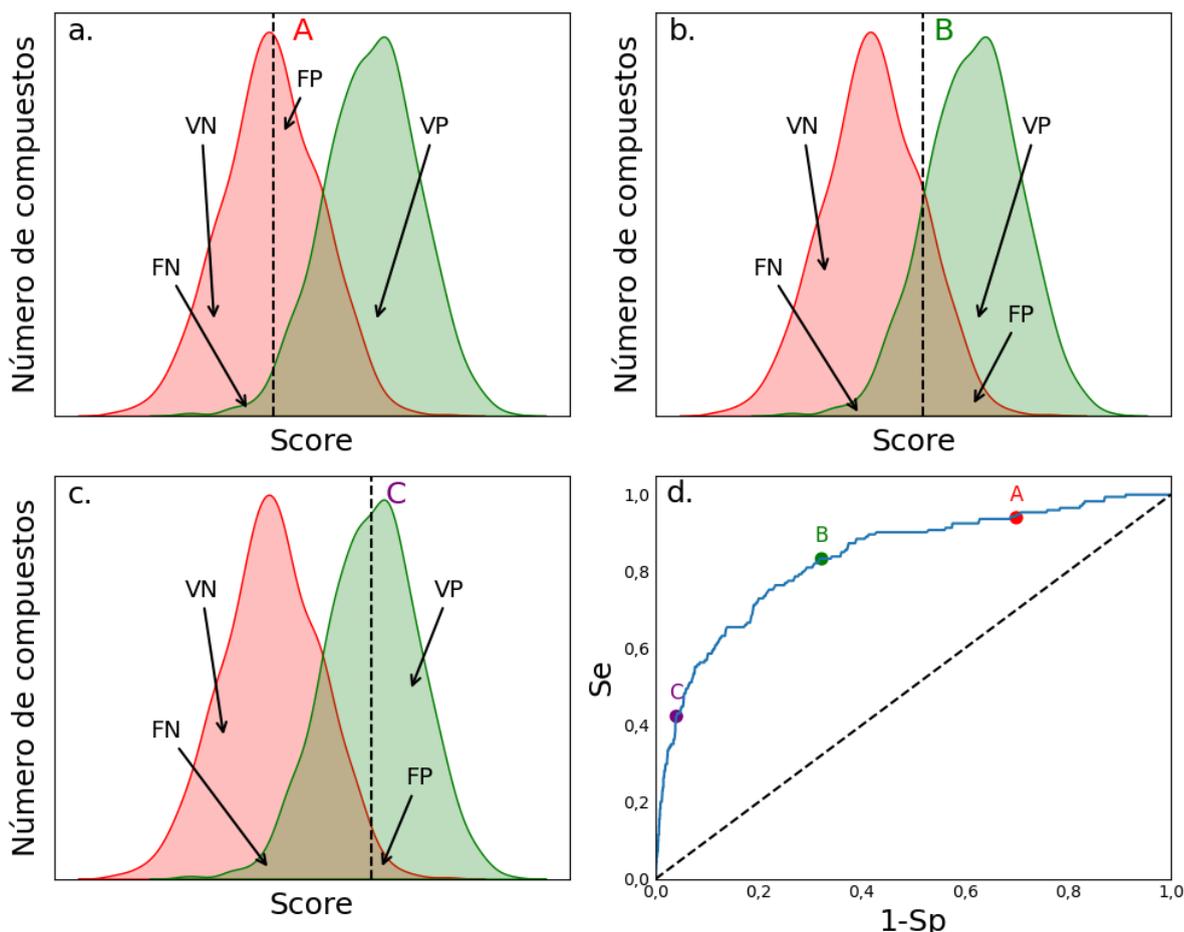


Figura 3.3.3. Esquema para la construcción de una curva **ROC**. Las gráficas a), b) y c) muestran el número de compuestos en el eje y y el **score** obtenido por el modelo en el eje x. Estas gráficas permiten ver cómo cambian las proporciones de **VN**, **VP**, **FN** y **FP** al modificar el valor de corte del **score** en los puntos A, B y C (líneas punteadas). Por último, la curva **ROC** se muestra en la figura d). Adaptado de Alberca, Tesis doctoral (2018).

3.3.4 CAMPAÑAS SIMULADAS DE CRIBADO VIRTUAL

Truchon y Bayly (2007) han demostrado, mediante experimentos de **CV** retrospectivos, que la métrica de enriquecimiento **AUCROC** depende de la proporción de compuestos activos e inactivos en el conjunto de datos a evaluar. La desviación estándar de la métrica converge a un valor constante al utilizar pequeñas proporciones de compuestos activos (también conocido como rendimiento de activos o prevalencia de activos, **Ya**). Se ha demostrado que valores de **Ya** por debajo de 0,05 proporcionan resultados más robustos.

Es importante tener en cuenta que los modelos obtenidos se aplicarán a campañas de **CV**, en las cuales se espera que haya un pequeño número de compuestos activos inmersos en un gran número de compuestos inactivos. Por lo tanto, es deseable que los modelos sean evaluados en una situación similar a la real. Para

evaluar el rendimiento de los modelos en campañas simuladas de **CV**, se dispersaron los compuestos activos que no forman parte del conjunto de entrenamiento entre una gran cantidad de compuestos presuntamente inactivos, conocidos como señuelos o *decoys*, ya mencionados en el presente capítulo. Estos señuelos fueron generados mediante dos métodos diferentes, los cuales se describen a continuación.

3.3.4.1 QUIMIOTECA SIMULADA LUDE

Para generar el primer grupo de *decoys*, se ha utilizado la herramienta de generación de señuelos automática *LIDeB's useful decoys (LUDe)* (Prada Gori et al., 2022b). Con esta herramienta, es posible obtener compuestos que sean fisicoquímicamente similares a los compuestos de interés, pero topológicamente diferentes. En este caso, se buscaron moléculas con características fisicoquímicas similares a los compuestos **ACTIVOS** del conjunto de prueba, pero que al mismo tiempo fueran topológicamente diferentes, y por ende presumiblemente inactivas. La similitud en las propiedades fisicoquímicas entre las moléculas señuelo y los compuestos del conjunto de prueba favorece que la tarea clasificadora del modelo no sea trivial, mientras que las diferencias en la topología se requieren para minimizar la probabilidad de que los *decoys* se unan al mismo sitio de unión que el compuesto **ACTIVO**.

Brevemente, la generación de señuelos por **LUDe** se puede resumir en los siguientes puntos:

- Se proporcionan los compuestos activos en notación **SMILES** (Weininger, 1988) al servidor de **LUDe**. Los compuestos son estandarizados con el paquete MolVS (Swain, 2019), como ya se describió anteriormente.
- Se calculan, para cada activo estandarizado, el peso molecular (**MW**), logaritmo del coeficiente de partición octanol-agua (**Log P**), número de enlaces rotables (**nRotB**), número de átomos aceptores de hidrógeno (**nHAcc**), número de átomos donores de hidrógeno (**nHDon**) y carga formal (**Charge**), utilizando el módulo Chem del paquete rdkit (Rational Discovery LLC, 2006).
- Los señuelos para cada activo se buscan en la base de datos ChEMBL30 utilizando los siguientes valores de corte para cada uno de los descriptores calculados en el paso anterior: **MW** \pm 20 Daltons, **LogP** \pm 0,5 unidades logarítmicas, **nRotB** \pm 1 enlace, **nHAcc** \pm 1 enlace, **nHDon** \pm 1 enlace, y **Charge** \pm 1 unidad.
- Luego, se aplican tres filtros topológicos sucesivos para seleccionar aquellas moléculas más disímiles topológicamente. En primer lugar, se aplica el coeficiente de Tanimoto (Willett, 2006) entre cada compuesto activo y sus respectivos potenciales *decoys*, tomando aquellos cuyo coeficiente sea menor a 0,2. Luego, se calcula la máxima subestructura común (**MCS**) entre

cada activo y sus potenciales señuelos y se calcula la fracción de átomos presentes en la **MCS (fMCS)** (Yang et al., 2010), con respecto a cada activo. Los *decoys* cuyo **fMCS** sea menor a 0,5 serán retenidos. Finalmente, se compara el *scaffold* de Murcko (Bemis & Murcko, 1996) entre cada activo y sus potenciales señuelos, reteniendo aquellos en los que dicho *scaffold* sea diferente del correspondiente en el compuesto activo.

- Por último, se comparan los *decoys* obtenidos contra todos los activos, reteniendo únicamente aquellos con un coeficiente de Tanimoto menor a 0,2, a fin de asegurar diferencia topológica entre todos los señuelos y todos los activos proporcionados por el usuario.

3.3.4.2 QUIMIOTECA SIMULADA DUDE-Z

Para generar el segundo conjunto de señuelos, se utilizó la versión optimizada de la herramienta *Enhanced Directory of Useful Decoys (DUDE-Z)* (Stein et al., 2021). De la misma manera que **LUDe**, esta herramienta permite obtener compuestos que son similares fisicoquímicamente a los compuestos deseados, pero con diferencias topológicas. En esencia, el algoritmo de generación de *decoys* en **DUDE-Z** es muy similar a su predecesor, **DUD-E**, aunque con algunas mejoras en la etapa de comparación entre los potenciales señuelos y los ligandos en cuanto a las cargas y estados de protonación, evitando el desbalance de cargas entre los ligandos y los *decoys* (Stein et al., 2021).

En resumen, la herramienta **DUDE-Z** realiza las siguientes acciones:

- Se compara el compuesto activo deseado frente a un pool de compuestos de la base de datos **ZINC15** (Sterling & Irwin, 2015) utilizando el **MW**, el **Log P** estimado teóricamente por el método de Moriguchi (**mLogP**), el **nRotB**, **Charge**, el **nHDon** y **nHAcc**.
- Se calcula la similitud molecular mediante *fingerprints* **ECFP4** (acrónimo de *Extended Connectivity Fingerprint 4*) (Rogers & Hahn, 2010) entre el compuesto activo y los potenciales señuelos encontrados.
- Se retienen solo los *decoys* que tienen similitud menor a 0,35 con respecto al compuesto activo. Si hay más de 50 compuestos, se retienen únicamente los 50 más disímiles.

3.4 ENSAMBLE DE MODELOS

El aprendizaje por ensamblado, también conocido en inglés como *ensemble learning*, utiliza múltiples algoritmos de aprendizaje para mejorar la capacidad predictiva en comparación con la utilización individual de cualquiera de ellos (Li et al., 2014). Los clasificadores por ensamble, que son modelos combinados o meta-clasificadores, han demostrado proporcionar una mejor generalización y

precisión que los clasificadores individuales (Daho & Chikh, 2015; Min, 2016; Carbonneau et al., 2016). Por lo tanto, pueden ser particularmente útiles para evitar el sobreajuste en situaciones donde el conjunto de datos de entrenamiento es pequeño en comparación con el número de posibles descriptores moleculares, un escenario común en el diseño de fármacos (Vyskovsky et al., 2016).

Para realizar los ensambles de modelos, se han utilizado cinco esquemas distintos de combinación de puntajes o *scores*:

1. **PRODUCTO (PROD)**: Este operador calcula el producto entre los puntajes individuales de los modelos combinados.
2. **MÍNIMO (MIN)**: Retiene el puntaje más bajo entre los puntajes individuales de los modelos combinados.
3. **PROMEDIO (PROM)**: Este operador calcula el promedio de los puntajes individuales de los modelos combinados.
4. **PROMEDIO DEL RANKING (RANK)**: Este esquema calcula el promedio del ranking de cada compuesto en cada modelo individual. Para ello, se ordenan los compuestos según su puntaje en cada modelo y se les asigna un número de ranking (del 1 al número total de compuestos). Luego, se promedian los rankings de los modelos individuales combinados.
5. **PROMEDIO DEL VOTO (VOT)**: El voto se calcula de acuerdo con la ecuación previamente utilizada por Zhang y Muegge (2006). Para cada compuesto, el voto para un modelo específico se calcula mediante la **Ecuación 3.4**.

$$voto_j = \max\left(0, \text{entero}\left(11 - \frac{rank_j}{0,02N_{BD}}\right)\right)$$

(3.4)

En este proceso, la posición en el ranking que ocupa cada compuesto j en un modelo determinado se indica mediante $rank_j$. Los compuestos se ordenan del mejor al peor puntaje en ese modelo, N_{BD} representa el número total de compuestos en la base de datos utilizada. La función *entero* redondea un número al entero inferior más cercano. Este procedimiento asigna 10 votos para el primer 2% de los compuestos, 9 votos para el siguiente 2%, y así sucesivamente hasta el 18%. A partir de ahí, solo se otorga un voto para los compuestos que se encuentran en el rango entre el 18% y el 20%, y los compuestos clasificados en el 80% restante no reciben votos. Para obtener el **VOT**, se promedian los votos de los modelos individuales.

Por otra parte, para la evaluación del desempeño de cada uno de los ensambles y elección del ensamble óptimo se utilizó la métrica **BEDROC**. Del acrónimo anglosajón *Boltzmann-Enhanced Discrimination of Receiver Operating*

Characteristics, **BEDROC** permite la evaluación del enriquecimiento temprano en las campañas de **CV**. Primeramente, los compuestos se rankean de acuerdo al *score* del modelo obtenido. Seguidamente, el valor de la métrica se asigna según una función exponencial decreciente que depende de la posición que tenga cada compuesto ACTIVO en la lista ordenada (r_i), como se observa en la **Ecuación 3.5**, donde $Atot$ se corresponde con el número total de ACTIVOS y α con el factor exponencial. En contraposición con el **AUCROC**, que refleja la posición promedio de los activos en el ranking del modelo, **BEDROC** pondera de manera preferencial los compuestos en las primeras posiciones del ranking. Esta métrica adopta valores entre 0 y 1, siendo la probabilidad de que un compuesto ACTIVO clasificado por el modelo evaluado se encuentre dentro de las primeras posiciones del ranking de la lista ordenada. Según la **Ecuación 3.5**, el término exponencial está gobernado por el parámetro α . Valores mayores de α hacen a un decaimiento más rápido de la curva, por que el valor que asume la métrica dependerá muy especialmente de la presencia de compuestos activos en las primeras posiciones del ranking. Específicamente, el grueso del valor de la métrica es dado por los ACTIVOS localizados en el $1/\alpha$ % de los compuestos mejor rankeados de la quimioteca sometida a **CV**.

$$\begin{aligned}
 BEDROC = & \left(\frac{\sum_{i=1}^{Atot} e^{-\alpha r_i / N}}{\frac{Atot}{N} \frac{1-e^{-\alpha}}{e^{(\alpha/N)}-1}} \right) \left(\frac{\frac{Atot}{N} \sinh \frac{\alpha}{2}}{\cosh \frac{\alpha}{2} - \cosh \left(\frac{\alpha}{2} - \alpha \frac{Atot}{N} \right)} \right) \\
 & + \frac{1}{1-e^{\alpha \left(1 - \frac{Atot}{N} \right)}} \quad (3.5)
 \end{aligned}$$

Al ser α un parámetro exponencial, no es posible comparar valores de **BEDROC** a diferentes valores de α . También es importante aclarar que valores de **BEDROC** de 0,5 son considerados favorables, a diferencia de las curvas **ROC**, debido al carácter exponencial de la primera (Truchon & Bayly, 2007).

En este trabajo de tesis se utilizó un valor de $\alpha=100$, lo que indica que el 1,6% inicial de la lista ordenada contribuye al 80% del valor de la curva **BEDROC** (Truchon & Bayly, 2007).

Para llevar adelante el ensamble de modelos, se tomaron los 4000 modelos individuales generados y se los ordenó de mayor a menor en términos de **AUCROC**. Luego, se descartaron aquellos modelos que compartían más de un descriptor molecular, reteniendo aquel con mayor **AUCROC** del par. Finalmente, cada ensamble de modelos con su respectivo operador fue evaluado mediante la métrica

BEDROC, optando por el ensamble en el que dicha métrica presentara un valor máximo. El cálculo de la métrica **BEDROC** fue realizado con el módulo *Scoring* del paquete rdkit (Rational Discovery LLC, 2006).

3.5 SELECCIÓN DEL PUNTO DE CORTE PARA EL CRIBADO

En una campaña de **CV** es importante estimar la probabilidad con la que un compuesto *hit* pueda confirmar su actividad predicha en los ensayos experimentales. Dicha estimación puede llevarse adelante mediante la construcción y análisis de gráficos de superficie del Valor Predictivo Positivo (**PPV**, del inglés *Positive Predictive Value*). Sin embargo, el valor de **PPV** es imposible de calcular para las quimiotecas sometidas al **CV** prospectivo, ya que se desconoce el valor de **Ya**, tal como se muestra en la **Ecuación 3.6**:

$$PPV = \frac{Se*Ya}{Se*Ya + (1-Sp)*(1-Ya)} \quad (3.6)$$

donde, **Ya** se corresponde con la proporción de **ACTIVOS** en la quimioteca, mientras que **Se** y **Sp** son la sensibilidad y especificidad, respectivamente.

Para construir la gráfica de superficie **PPV** se calcula la **Se** y **Sp** para todo el rango de valores de *score* posibles para el modelo y/o ensamble de modelos utilizando la un subconjunto de la quimioteca simulada **LUDe (LUDe-B)**, y se observa la relación entre el valor de **PPV** en función de la relación **Se/Sp**, variando **Ya** entre 0,01 y 0,001, que se corresponde con el rango de proporciones de **ACTIVOS** identificados habitualmente en un cribado farmacológico de alto rendimiento (Alberca et al., 2018; Morales et al., 2020).

3.6 CRIBADO VIRTUAL

Después de seleccionar los modelos con los mejores resultados, estos se aplicaron a ocho bases de datos para identificar posibles compuestos que pudieran inhibir la actividad de las **FABPs** de cestodos. Se calcularon los descriptores moleculares y se obtuvieron los puntajes para cada compuesto en cada base de datos. A continuación, se utilizaron las siguientes bases de datos para el cribado:

- **Drugbank**: Es un repositorio en línea que almacena información sobre fármacos aprobados por la **FDA** para su uso clínico, medicamentos en proceso de evaluación clínica y compuestos en etapas iniciales de evaluación. En su totalidad, contiene 15550 pequeñas moléculas de tipo

fármaco, de las cuales 2742 son fármacos ya aprobados por la **FDA**, según lo reportado por Wishart et al. (2018). La versión 5.1.6 de esta quimioteca fue sometida a **CV**.

- **DRH**: Es una base de datos manualmente curada con fármacos aprobados por la **FDA**, como así también en ensayos clínicos y preclínicos. En su última versión disponible (Marzo 2020), este repositorio cuenta con un total de 6927 compuestos, de los cuales 2427 son compuestos presentes en el mercado estadounidense (Corsello et al., 2017). Dicha versión de **DRH** fue la seleccionada para **CV**.
- **Sweetlead**: Es una base de datos que incluye medicamentos aprobados por organismos reguladores de otros países, como Australia, China e India, así como compuestos extraídos de plantas medicinales tradicionales y otros productos químicos regulados. En total, esta base de datos contiene 4442 moléculas, de las cuales 2836 son fármacos aprobados (Novick et al., 2013). La versión 1.0 de esta quimioteca fue la utilizada para el cribado prospectivo.
- **HybridMolDB**: Esta base de datos incluye 4927 moléculas híbridas, entendiendo como tales a aquellas entidades químicas que resultan de la combinación de dos o más moléculas con bioactividad. Este tipo de compuestos pueden presentar mejoras en propiedades farmacológicas con respecto a sus moléculas constitutivas individuales, como por ejemplo poder interactuar con diversos blancos, reducir efectos no deseados o interacciones medicamentosas (Li et al., 2019). La versión 1.0 fue la utilizada en este trabajo.
- **NuBBEDB**: Esta base de datos contiene 2223 compuestos (Noviembre 2023, versión utilizada en esta tesis) derivados de productos naturales aislados de diferentes regiones de la República Federativa de Brasil, incluyendo información química, biológica, farmacológica y espectroscópica de cada uno de dichos compuestos (Valli et al., 2013; Pilon et al., 2017).
- **NaturAr**: Esta base de datos es desarrollada y mantenida por el grupo de Química Computacional del CEQUINOR (UNLP) y contiene alrededor de 700 compuestos (Noviembre 2023, versión sometida a **CV**) derivados de productos naturales cuyas especies son autóctonas en nuestro país (*Computational Chemistry Group*, 2023).
- **COCONUT**: Es una base de datos que compendia más de 400000 compuestos derivados de productos naturales, junto con datos de estereoquímica, estudios en literatura, organismos productores, distribución geográfica y propiedades moleculares (Sorokina et al., 2021). Su última actualización disponible (Enero 2022) fue la utilizada para el **CV**.
- **LIDeB**: Esta base de datos es un compendio de la quimioteca interna del laboratorio, es decir diversos compuestos adquiridos en el pasado y/o sintetizados en el contexto de otros proyectos de investigación. La misma cuenta con unos 150 compuestos.

3.7 DOMINIO DE APLICACIÓN

La construcción de modelos **QSAR** a partir de conjuntos de entrenamiento con poca diversidad molecular limita su aplicabilidad a quimiotecas externas de compuestos. Es fundamental establecer un dominio de aplicación (**DA**), también conocido como dominio de aplicabilidad, para estos modelos, con el fin de determinar el conjunto de estructuras químicas que pueden ser predichas de manera confiable. En resumen, la capacidad de generalización de un modelo predictivo depende, en parte, de la amplitud de su **DA**. Si el **DA** es demasiado restringido, implica que el modelo sólo puede proporcionar predicciones confiables para un número limitado de moléculas (Dimitrov et al., 2005).

Se han propuesto diversos enfoques para determinar el **DA**, los cuales varían en la forma en que caracterizan el espacio de interpolación basado en los descriptores utilizados (Sahigara et al., 2012).

En nuestro caso, hemos utilizado un método llamado "regla de la palanca" (*leverage* en inglés) para determinar qué compuestos de la base de datos de cribado se encuentran dentro del **DA**. Este método se basa en el cálculo de los valores de palanca (h_i) para los compuestos evaluados, donde x_i representa el vector de descriptores del compuesto i , y X es la matriz generada a partir de los valores de descriptores del conjunto de entrenamiento del modelo (Tropsha et al., 2003), tal como se muestra en la **Ecuación 3.7**.

$$h_i = x_i^T (X^T X)^{-1} x_i \quad (3.7)$$

Consecuentemente, un compuesto de la base de datos del cribado se considerará dentro del **DA** de un modelo si su valor de palanca (h_i) es inferior a un umbral establecido (h^*). En la presente tesis, el umbral fue definido como se indica en la **Ecuación 3.8** :

$$h^* = \frac{2f}{n} \quad (3.8)$$

donde f representa el número de descriptores incluidos en el modelo, y n es el número de compuestos que conforman el conjunto de entrenamiento.

3.8 MÉTODOS DIRECTOS. MODELADO ESTRUCTURAL DE LAS FABPs Y GENERACIÓN DE MODELOS DE DOCKING

Como ya se discutió en el capítulo 2, los métodos directos de **CV** se basan en el cálculo de interacciones moleculares entre cada una de las moléculas que forman parte de la quimioteca virtual y el blanco molecular del cual se desea modular su actividad biológica. En consecuencia, es menester contar con la estructura tridimensional de dicho blanco, ya sea de manera experimental o teórica. A lo largo de este apartado se detalla, por un lado la generación de los modelos tridimensionales de las diferentes isoformas de **FABPs** y posterior generación de ensamblajes conformacionales para cada una de ellas, y por otro, la utilización de dichos ensamblajes en la estrategia de **CV** por *docking* molecular, como así también la validación de estos modelos. En la **Figura 3.8**, se muestra un esquema general del procedimiento seguido en el presente trabajo de tesis doctoral.

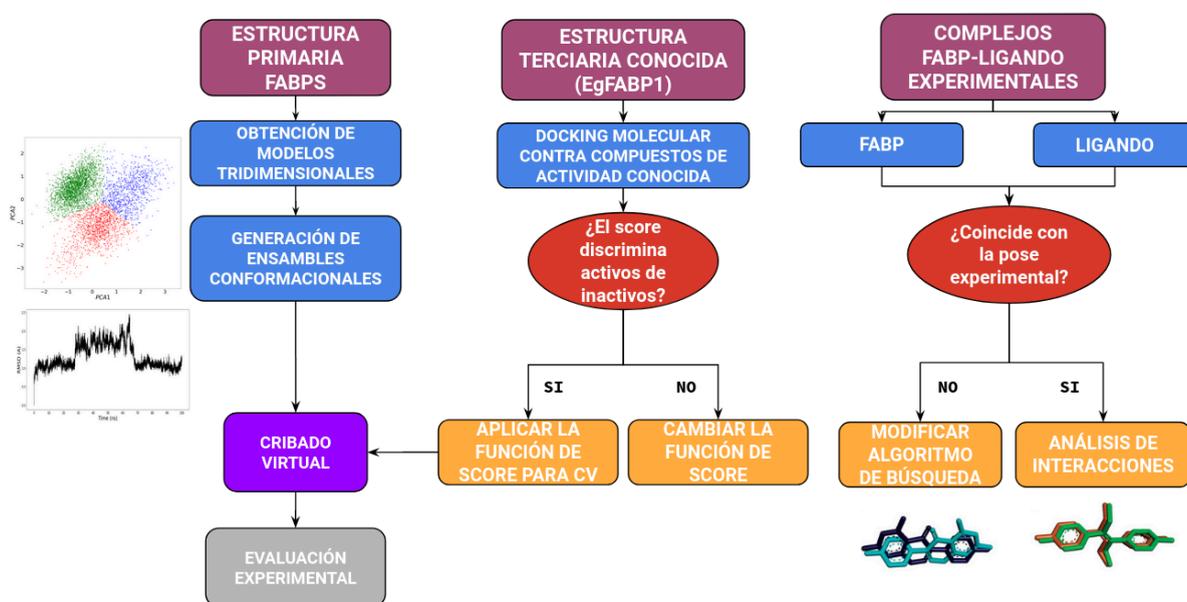


Figura 3.8. Flujo de trabajo utilizado para el **CV** basado en la estructura.

3.8.1 MODELADO ESTRUCTURAL DE LAS FABPs

En los años 70, Anfinsen evidenció que la estructura de una proteína es determinada por su secuencia de aminoácidos, sentando las bases de lo que hoy se conoce como predicción estructural de proteínas (Pearce & Zhang, 2021). A lo largo de estos 50 años de investigación se lograron avances que, a su vez, abrieron la puerta a métodos de predicción aún más potentes. Por ejemplo, el desarrollo del algoritmo AlphaFold2, ampliamente usado hoy en día, no hubiera sido posible sin los métodos predecesores, como los métodos *ab-initio* o el modelado por homología, los cuales a su vez no hubieran existido sin el desarrollo previo de metodología experimental para la determinación estructural de proteínas, cuyo

primer gran hito fue la obtención del cristal de la hemoglobina humana (Perutz et al., 1960).

Como ya se ha mencionado en el capítulo 1, las **FABPs** son una familia de proteínas que desempeñan un papel crucial en el transporte y la regulación del metabolismo de los ácidos grasos (Lee et al., 2021). Presentan una estructura terciaria caracterizada por un barril compuesto por hojas β , y un dominio hélice α minoritario sobre el extremo N terminal. Aunque la estructura básica es altamente conservada, las diferencias a nivel de secuencia primaria le dan a cada isoforma de **FABP** diferentes afinidades a distintos tipos de lípidos (Smathers & Petersen, 2011).

Tal como fue descrito en el primer capítulo del presente trabajo, *E. granulosus* expresa en total seis isoformas de **FABPs** (denominadas aquí **EgFABP1**, **EgFABP1.2**, **EgFABP2**, **EgFABP3**, **EgFABP4** y **EgFABP5**) (Pórfido et al., 2020), mientras que *E. multilocularis* hace lo propio con un total de cinco isoformas (llamadas **EmFABP1**, **EmFABP2**, **EmFABP3**, **EmFABP4** y **EmFABP5**) al igual que *T. solium* (**TsFABP1**, **TsFABP2**, **TsFABP3**, **TsFABP4** y **TsFABP5**) (WormBase ParaSite, *Taenia solium*, 2023).

En los siguientes apartados se describe el modelado estructural de quince de las mencionadas isoformas, ya que **EgFABP1** cuenta con estructura determinada experimentalmente (PDB:1O8V). El modelado estructural planteado en esta tesis consiste en el compilado de las secuencias de las isoformas de **FABPs** de cestodos, la generación de los modelos estructurales en ColabFold (Mirdita et al., 2021), la generación de ensamblajes conformacionales para cada isoforma por **DM** y la posterior selección de dichas poses para el **CV** por *docking*.

3.8.1.1 COMPILADO DE LAS SECUENCIAS PARA CADA ISOFORMA

Las secuencias aminoacídicas para las isoformas de **FABPs** de *E. granulosus* (**EgFABP1.2**, **EgFABP2**, **EgFABP3**, **EgFABP4** y **EgFABP5**) y *E. multilocularis* (**EmFABP1**, **EmFABP2**, **EmFABP3**, **EmFABP4** y **EmFABP5**) fueron extraídas del trabajo de Pórfido y colaboradores (Pórfido et al., 2020), mientras que las isoformas presentes en *T. solium* (**TsFABP1**, **TsFABP2**, **TsFABP3**, **TsFABP4** y **TsFABP5**) se obtuvieron del sitio WormBase (WormBase ParaSite, *Taenia solium*, 2023). En el **Material Suplementario 3 (Tabla S3a)** se pueden encontrar cada una de las secuencias primarias de las mencionadas isoformas.

3.8.1.2 GENERACIÓN DE LAS ESTRUCTURAS TRIDIMENSIONALES

Cada una de las secuencias compiladas fue enviada al servidor de ColabFold (The AlphaFold Team & Google Colab, 2022) para el cómputo de la estructura

tridimensional correspondiente, utilizando los parámetros por defecto, obteniendo cinco modelos posibles por isoforma.

Cada modelo obtenido para cada isoforma de **FABP** fue evaluado mediante el cálculo del puntaje **LDDT** (del inglés *Local Distance Difference Test*). Este puntaje evalúa qué tan bien se reproducen las distancias interatómicas en una estructura modelada en comparación con al menos una proteína de referencia (Mariani et al., 2013). ColabFold clasifica los modelos generados según el puntaje **LDDT**, y proporciona también un puntaje **LDDT** por residuo (**pLDDT**), ambos en el rango entre 0-100. Valores mayores de **LDDT** se corresponden con predicciones estructurales más precisas. Para los propósitos de este trabajo, se consideró el modelo con mayor **LDDT** global para cada isoforma de **FABP**. Muy brevemente, el **score LDDT** genera una lista de distancias interatómicas tanto en de la proteína de referencia como de la estructura a comparar. Para cada átomo i , todos los átomos j pertenecientes a un aminoácido diferente al del átomo i que se encuentren a una distancia menor o igual a 5 Å en la estructura de referencia son considerados átomos interaccionantes de i . Esta lista de interacciones ij de la estructura de referencia son comparadas contra las distancias encontradas en los átomos correspondientes en el modelo predicho. Si la diferencia de distancias entre los átomos de la referencia y el modelo se encuentra dentro un valor umbral, la interacción se considera como preservada en la predicción. Se utilizan cuatro valores diferentes de umbral (0,5; 1; 2 y 4 Å) y el **score LDDT** se calcula entonces como el promedio de la fracción de interacciones preservadas en el modelo en cada uno de los valores umbral (Mariani et al., 2013).

El algoritmo de ColabFold predice el puntaje de **pLDDT**, junto con el modelo estructural. ColabFold está entrenado con valores de **pLDDT** verdaderos, es decir calculados para predicciones de proteínas resueltas experimentalmente. De esta manera, el **score pLDDT** es predicho a lo largo de la estructura proteica, mientras que el **LDDT** es el promedio de los **pLDDT** a lo largo de la secuencia de aminoácidos, estando la/las estructura/as de referencia dada/s por el conjunto de entrenamiento implementado en el algoritmo (Jumper et al., 2021, Material Suplementario).

3.8.1.3 GENERACIÓN DE ENSAMBLES CONFORMACIONALES POR DINÁMICA MOLECULAR

La **DM**, ya descrita en el capítulo 2, es una técnica computacional utilizada para simular el movimiento y la interacción de átomos, que permite, entre muchas otras cosas, explorar y buscar ensambles conformacionales de la molécula del receptor (Amaro et al., 2018).

En el contexto de este trabajo, la **DM** se utiliza para generar ensambles conformacionales de la proteína receptora, explorando diferentes posiciones y

orientaciones relativas entre ellos. Estos ensamblajes conformacionales se pueden utilizar posteriormente en campañas de *docking* para poder determinar la mejor conformación de unión y la afinidad entre cada ligando y la proteína receptora, estrategia conocida en inglés como *ensemble docking* (Amaro et al., 2018).

Las simulaciones de **DM** se realizaron utilizando el programa GROMACS 2020.4 (Berendsen et al., 1995; Lindahl et al., 2001; Van Der Spoel et al., 2005; Hess et al., 2008; Pronk et al., 2013; Páll et al., 2015; Abraham et al., 2015), usando el campo de fuerza OPLS-AA (Jorgensen et al., 1988). Las proteínas fueron solvatadas con 6108 moléculas de agua OPC (Izardi et al., 2014) en una caja dodecaédrica de 205,13 nm³, llevando el sistema a neutralidad con contraiones.

Posteriormente, la energía de cada sistema se minimizó utilizando el algoritmo *steepest descent* hasta alcanzar un valor máximo de fuerza inferior a 1000 kJ mol⁻¹nm⁻¹, con un tamaño de paso (*step size*) de 0,01. En segundo lugar, se llevó a cabo para cada isoforma un ensamble **NVT** (es decir, número de partículas, volumen y temperatura constantes) de 100 ps con restricciones de posición para los átomos de la proteína a 300 K de temperatura, utilizando el termostato modificado de Berendsen (Bussi et al., 2007) con una constante de acoplamiento de temperatura (τ_t) igual a 0,1 ps, y con un tiempo de integración (*time step*) de 4 fs.

Seguidamente, se realizó un ensamble **NPT** (número de partículas, presión y temperatura constantes) de 100 ps con restricciones de posición utilizando el barostato de Parrinello-Rahman (Parrinello & Rahman, 1981), estableciendo la presión en 1 bar con una constante de acoplamiento de la presión (τ_p) igual a 2 ps, un coeficiente de compresibilidad isotérmico (γ) de 4,5x10⁻⁵ bar⁻¹ y un tiempo de integración, también, de 4 fs.

Finalmente, se llevó a cabo una producción de **DM** sin restricciones durante 100 ns para cada isoforma en las mismas condiciones mencionadas anteriormente para el ensamble **NPT**.

3.8.1.4 SELECCIÓN DE CONFORMACIONES

Una vez generadas las trayectorias de cada isoforma de las **FABPs**, es necesario contar con herramientas de análisis a fin de poder extraer información útil de cada una de ellas.

Para ello, sobre cada trayectoria generada se llevó a cabo el análisis de **PCA**. En el contexto de la **DM**, **PCA** se aplica para analizar y visualizar las fluctuaciones y correlaciones de los movimientos de las partículas en un sistema molecular (Palma & Pierdominici-Sottile, 2023).

El **PCA** en **DM** se basa en el cálculo de las matrices de covarianza o correlación de las trayectorias de la **DM**. Estas matrices capturan la variabilidad y las relaciones

lineales entre las coordenadas atómicas a lo largo del tiempo. Al aplicar **PCA** a estas matrices, se obtienen las componentes principales, que son combinaciones lineales de las coordenadas originales que explican la mayor parte de la variabilidad en los datos (Palma & Pierdominici-Sottile, 2023).

Una vez obtenidas las dos primeras componentes principales para cada trayectoria, se aplicó el algoritmo *k-means* (Böcker, 2008), variando sistemáticamente el número de clusters (**k**) desde 2 hasta 20. De la misma manera, se variaron los valores de la semilla entre 2 a 20, junto con el número de clusters. Los *clusters* se evalúan calculando el puntaje **SIL** (Rousseeuw, 1987) para cada valor de **k**, seleccionando el valor de **k** cuyo **SIL** sea el más alto (al que llamamos k-máximo). Cabe aclarar, que en este caso, cada uno de los puntos agrupados por *k-means* se corresponde con una foto o *snapshot* de la simulación. Por lo tanto, el valor de **k** representa el número de grupos, en donde cada una de estas *snapshot* será asignada a un *cluster* determinado, de acuerdo a sus coordenadas de **PCA**. De esta manera, se extrajeron para cada isoforma cada uno de los centroides correspondientes con el valor de k-máximo respectivo. Los residuos elegidos para cada isoforma en el cálculo de la matriz de covarianza se indican en la **Tabla 3.8.1.4**. Tres residuos (para **EgFABP1.2**, **EgFABP2**, **EgFABP5**, **EmFABP1**, **EmFABP2**, **EmFABP5**, **TsFABP1** y **TsFABP2**: 107-127-129; para **EgFABP3**, **EmFABP3** y **TsFABP3**: 116-137-139; **TsFABP5**: 109-129-131) se corresponden con la tríada de las **FABPs** (descrita en la siguiente sección) mientras que el resto de los residuos fueron seleccionados de acuerdo a estudios previos de dinámica molecular, donde postulan a dichos residuos como importantes en la interacción con la cola hidrofóbica de ligandos de las **FABPs** (Jakobsson et al., 2003). El análisis de **PCA** fue realizado en GROMACS y tanto el cálculo de *k-means* como de **SIL** fue realizado por la biblioteca de Python scikit-learn (Pedregosa et al., 2011).

Tabla 3.8.1.4. Residuos utilizados en el cálculo de la matriz de covarianza, para cada isoforma de las **FABPs**.

<u><i>Isoforma</i></u>	<u><i>Aminoácidos del sitio de unión</i></u>
<i>EgFABP1.2</i>	16-76-77-107-127-129
<i>EgFABP2</i>	16-76-77-107-127-129
<i>EgFABP3</i>	16-80-81-116-137-139
<i>EgFABP5</i>	16-76-77-107-127-129
<i>EmFABP1</i>	16-76-77-107-127-129
<i>EmFABP2</i>	16-76-77-107-127-129

<i>EmFABP3</i>	16-80-81-116-137-139
<i>EmFABP5</i>	16-76-77-107-127-129
<i>TsFABP1</i>	16-76-77-107-127-129
<i>TsFABP2</i>	16-76-77-107-127-129
<i>TsFABP3</i>	16-80-81-116-137-139
<i>TsFABP5</i>	16-76-77-109-129-131

3.8.1.5 OBTENCIÓN DE SITIOS DE HIDRATACIÓN POR DINÁMICA MOLECULAR

Es sabido que las diferentes isoformas de **FABPs** de cestodos presentan tres residuos altamente conservados en sus respectivos sitios de unión, los cuales tienen un rol preponderante en el proceso de unión de ligandos (Jakobsson et al., 2003): Arg107, Arg127 y Tyr129. Estos residuos son conocidos como 'la tríada de las **FABPs**'.

La estructura cristalina de **EgFABP1** (PDB:1O8V), al presente la única estructura tridimensional resuelta de **FABPs** de cestodos, se encuentra co-cristalizada con Ácido palmítico, cuyo grupo carboxilato interacciona con la mencionada tríada, como así también con moléculas de agua que merodean el sitio de unión y se cree que son importantes para el proceso de unión del ligando (Jakobsson et al., 2003).

Con este último punto en mente, se decidió determinar sitios de hidratación en el sitio de unión de **EgFABP1** mediante **DM**. Para ello, se llevó adelante una simulación de **DM**, de la misma manera que se explicó en la sección 3.8.1.3, utilizando la estructura 1O8V y removiendo la molécula de Ácido palmítico. Luego, los sitios de hidratación se determinaron utilizando WATCLUST (López et al., 2015), un algoritmo que detecta sitios de hidratación mediante *clustering* de moléculas de agua. Este algoritmo se encuentra disponible como paquete en **VMD** (acrónimo de *Visual Molecular Dynamics*) (Humphrey et al., 1996).

El algoritmo de WATCLUST permite calcular propiedades estructurales y termodinámicas de las moléculas de agua de solvatación. A partir de una simulación de **DM** de una proteína en un solvente explícito, el software es capaz de encontrar regiones del espacio donde hay una alta probabilidad de encontrar moléculas de agua, y calcular propiedades termodinámicas relevantes para diferentes análisis biofísicos (López et al., 2015). Aquellos sitios donde la probabilidad de encontrar

una molécula de agua es mayor que en el entorno se denominan sitios de hidratación (**WS**, *Water site* en inglés).

WATCLUST permite transferir la información de los **WS** al programa de *docking* Autodock-GPU para introducir sesgos en interacciones entre ligando y receptor (Arcon et al., 2019). La posibilidad de introducir sesgos en el modelo de *docking* permite puntuar de manera diferencial ciertas interacciones que podrían mejorar el desempeño de la predicción del modelo.

A continuación, la **Figura 3.8.1.5** esquematiza el funcionamiento del algoritmo de WATCLUST.

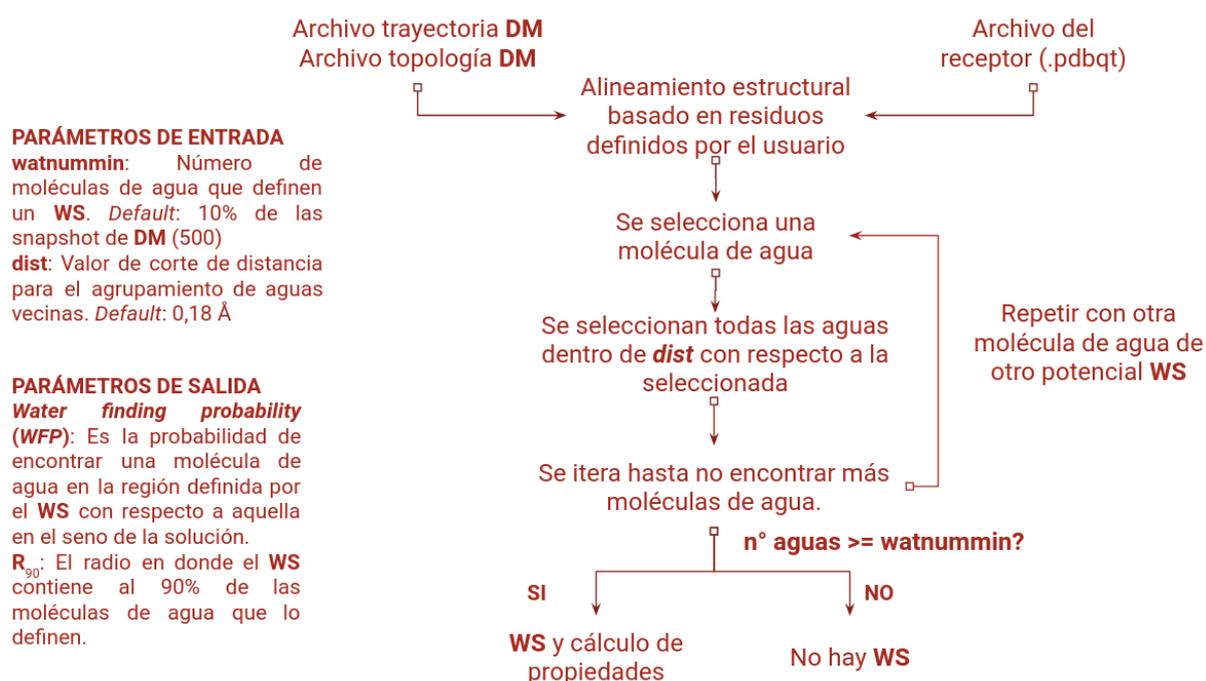


Figura 3.8.1.5. Esquema de funcionamiento del algoritmo WATCLUST.

3.8.2 GENERACIÓN DE MODELOS DE DOCKING

Como se discutió en el capítulo 2, el **CV** por *docking* es una técnica computacional utilizada para predecir la afinidad y el modo de unión entre un compuesto químico y un receptor. El objetivo principal del **CV** por *docking* es identificar y seleccionar compuestos prometedores que puedan unirse de manera específica a una proteína diana, siendo útil tanto para el diseño como para el descubrimiento de fármacos.

En la presente tesis doctoral se utilizaron programas de *docking* basados en dos métodos de búsqueda conformacional, y se evaluaron sus respectivos desempeños en términos de capacidad predictiva sobre un conjunto de datos de compuestos **ACTIVOS** e **INACTIVOS** conocidos, como así también en términos de predicción del

modo de unión. Para el primer caso, se utilizaron los compuestos **ACTIVOS** e **INACTIVOS** reportados en literatura contra **AFABP**, como así también los *decoys* generados por las herramientas **LUDe** y **DUDE-Z**. Para la evaluación del modo de unión, se utilizaron complejos **FABP**-ligando experimentales. A continuación, se hace una descripción de ambos métodos de búsqueda conformacional:

- **Algoritmo Genético Lamarckiano (LGA)**: Este tipo de algoritmos adaptan los conceptos de la teoría evolutiva a la resolución de problemas de optimización numérica. La idea fundamental es la generación de una población inicial de posibles soluciones al problema sobre las que se realizan sucesivas iteraciones dando lugar a una selección competitiva que mimetiza el proceso de selección natural, eliminando aquellas soluciones con valor de *score* desfavorable (Gavernet, 2021).

Aplicado a *docking*, cada pose del ligando se representa mediante un vector de genes, en donde cada gen es una variable que representa la posición, orientación y valores que asumen los ángulos de torsión del ligando.

El algoritmo genera una población inicial de conformaciones de partida al azar, donde cada una de ellas es dockeada en el sitio de unión de interés (Gavernet, 2021). Luego de ello, el algoritmo **LGA** selecciona en cada iteración, por un lado, las poses energéticamente más favorables según su función de *score* utilizando un esquema de selección conocido en inglés como *binary tournament* y, por otro lado, selecciona al azar un subconjunto de poses (6% por defecto) que serán optimizadas mediante un algoritmo de búsqueda local. El esquema de selección *binary tournament* (Santos-Martins et al., 2021, Material Suplementario) consiste en la selección al azar de las poses generadas en grupos de a dos, donde la pose cuyo *score* sea el más favorable del par será la seleccionada. Finalmente, las poses elegidas por *binary tournament* y las optimizadas por el algoritmo de búsqueda local son el punto de partida para la generación de nuevas poses mediante la recombinación y mutación de sus “genes”, las cuales serán dockeadas nuevamente frente al blanco molecular. Las iteraciones del algoritmo de **LGA** se detienen cuando se alcanza el número máximo de evaluaciones del *score* o el número máximo de generaciones, lo que ocurra primero. La pose con el mejor *score* es el resultado del **LGA** (Santos-Martins et al., 2021).

- **Algoritmo de Broyden-Fletcher-Goldfarb-Shanno (BFGS)**: Este algoritmo busca conformaciones de mínima energía mediante la minimización del vector gradiente de la función de puntuación, es decir, el conjunto de las derivadas parciales de la función de puntuación con respecto a la posición y orientación del ligando, expresado como vector. El número de pasos en una ejecución se determina de manera adaptativa, de acuerdo con la aparente complejidad del problema, y se realizan varias ejecuciones a partir de conformaciones aleatorias (Trott & Olson, 2009).

Se ensayaron un total de cuatro programas diferentes para la búsqueda conformacional, tres de ellos basados en **LGA** (AutoDock-GPU (**AD GPU**) (Santos-Martins et al., 2021), AutoDock-Hydrated (**AD Hyd**) (Forli & Olson, 2012) y AutoDock Bias (**AD Bias**) (Arcon et al., 2019)) y uno basado en **BFGS** (AutoDock Vina (**Vina**) (Trott & Olson, 2009), utilizando dos funciones de puntuación diferentes, la función nativa de **Vina** (Trott & Olson, 2009) y su versión optimizada **Vinardo** (Quiroga & Villarreal, 2016)), cuyas características se describen a continuación:

- **AD GPU:** Describe las interacciones ligando-receptor utilizando una función de puntuación basada en la física que ha demostrado ser útil en una variedad de proyectos de descubrimiento de fármacos. Esta versión permite la utilización de placas gráficas de video (en inglés *Graphic Processor Unit*, **GPU**), reduciendo drásticamente el tiempo de ejecución del *docking* con respecto a versiones previas ejecutadas en núcleos **CPU** (*Central Processor Unit*) (Santos-Martins et al., 2021).
- **AD Hyd:** Es una versión de **AD GPU** que tiene en cuenta potenciales **WS** en cada uno de los ligandos a cribar. Debido a que la representación y mediación de las moléculas de agua son muy importantes para las interacciones ligando-proteína, **AD Hyd** agrega moléculas de agua en los potenciales **WS** de cada uno de los ligandos a dockear y predice de manera automática cuál/es de dichas moléculas de agua son importantes para la unión al sitio activo. La función de *score* tiene en cuenta las contribuciones entálpicas y entrópicas de las aguas discretas en la unión del ligando, mejorando la precisión de la estimación de energía y el rendimiento del *docking*. El método no presupone ningún conocimiento previo sobre el estado de hidratación de la proteína o el ligando (Forli & Olson, 2012).
- **AD Bias:** Esta otra versión de **AD GPU**, es un método general que permite guiar los cálculos de *docking* hacia la formación de cualquier interacción proteína-ligando definida por el usuario mediante la aplicación de un sesgo (*bias* en inglés) sobre cierto tipo de átomo en un lugar dado de la grilla de *docking*. Estos sesgos consisten en una recompensa energética para el tipo de interacción que se desea en la región del espacio definida por el operador. Se pueden introducir varios sesgos para un tipo de átomo dado, como así también diferentes tipos de sesgo (donadores/aceptores de enlaces de hidrógeno, aromáticos o definidos por el usuario) simultáneamente (Arcon et al., 2019).
- **Vina:** Este método, como ya se mencionó, utiliza el algoritmo de búsqueda de **BFGS**. La función de *score* de **Vina** carece de un significado físico en sí mismo, consistiendo en un potencial similar a las fuerzas de Van der Waals (definido por una combinación de un término de repulsión y dos términos atractivos gaussianos), un término de enlace de hidrógeno no direccional, un término hidrofóbico y una penalización por entropía conformacional (Trott & Olson, 2009; Quiroga & Villarreal, 2016). En este trabajo de tesis se ha utilizado la versión en **CPU** de **Vina** (Eberhardt et al., 2021), aunque una versión en **GPU** ha sido publicada recientemente (Tang et al., 2022).

- ***Vinardo***: Es una función de puntuación que comparte términos de componentes con la función de puntuación de **Vina**, tales como interacciones estéricas, interacciones hidrofóbicas y enlaces de hidrógeno no direccionales, pero por otra parte **Vinardo** presenta un término de interacción estérica modificado, nuevos radios atómicos e interacciones simplificadas. (Quiroga & Villarreal, 2016).

En el **Material Suplementario 3 (Tabla S3b y Ecuaciones S3.1-S3.4)** se presentan las funciones de puntuación de cada uno de estos programas de *docking*, como así también los parámetros de corrida utilizados.

3.8.2.1 COMPILADO DE LA BASE DE DATOS

Los métodos de *docking* molecular, son utilizados en esta tesis doctoral con dos propósitos, por un lado, como predictor categórico de compuestos químicos y por otro, como predictor del modo de unión entre los diferentes ligandos y el receptor de interés. En consecuencia, y a diferencia de los métodos indirectos, se necesitan al menos dos tipos de bases de datos diferentes, una de ellas capaz de validar la función de puntuación del programa de *docking* (*scoring power* en inglés) y la otra capaz de validar la reproducción de la pose experimental (*docking power*) (Llanos et al., 2020; Llanos et al., 2021; Wang & Zhang, 2017).

Para la validación del *scoring power*, al igual que en los métodos basados en el ligando, es necesario contar con un conjunto de datos de compuestos químicos con actividad experimental reportada frente a nuestro blanco molecular. Sin embargo, a diferencia de los métodos indirectos, en la validación del *scoring power* no es necesaria la partición del conjunto de datos en una base de entrenamiento y otra de prueba, ya que la función de puntaje del *docking* es inherente al programa que se esté utilizando (Llanos et al., 2020; Llanos et al., 2021; Wang & Zhang, 2017). En esta validación se utilizaron los mismos compuestos compilados para la generación de los modelos desde el ligando, junto con ambas quimiotecas de señuelos (**LUDe** y **DUDE-Z**) (Prada Gori et al., 2022b; Stein et al., 2021).

Por otra parte, para la validación del *docking power* es necesario contar con un conjunto de complejos **FABP**-ligando resueltos experimentalmente. En este tipo de validaciones, se modifica el archivo de extensión pdb de cada complejo, en donde se separa por un lado el receptor y por otro lado, el ligando, y se procede a dockear. El algoritmo de *docking* entregará como archivo de salida un conjunto de poses predichas entre el receptor y el ligando. Para evaluar si una pose dada predice correctamente el modo de unión experimental, se computa entre ellas el **RMSD**, métrica que permite medir la distancia promedio entre los átomos de dos poses superpuestas, considerando como adecuado un **RMSD** menor a 2 Å para un *re-docking* y un **RMSD** menor a 3 Å en el caso de un *cross-docking* (Kirchmair, 2008). El **RMSD** es calculado mediante la **Ecuación 3.9**, donde x_i refiere a la

posición del átomo i en la predicción, x_{ref} es la posición del átomo i en la pose experimental y N es el número de átomos evaluados.

En esta validación se utilizaron 33 complejos **FABP**-ligando diferentes, en donde cada uno de los ligandos es diferente mientras que las estructuras de las **FABPs** se corresponden con diferentes isoformas humanas y de otras especies. Con esta selección se buscó ampliar la diversidad de ligandos y de sitios de unión posibles en las **FABPs**, a fin de poder establecer relaciones más generales en cuanto a los alcances en la predicción del modo de unión de los ligandos. En el **Material Suplementario 3 (Tabla S3c)** se compendian las estructuras tenidas en cuenta en esta validación.

$$RMSD = \sqrt{\frac{\sum_{i=1}^N (x_i - x_{ref})^2}{N}} \quad (3.9)$$

3.8.2.2 PREPARACIÓN DE LOS LIGANDOS

Típicamente, los programas de *docking* requieren que tanto el receptor como los ligandos se encuentren en formato de archivo con extensión `pdbqt`, que contiene, además de las coordenadas atómicas, la carga y el tipo de átomo para cada partícula.

Tanto los compuestos pertenecientes al conjunto de validación del *scoring power*, como aquellos pertenecientes al conjunto de validación del *docking power* y aquellos pertenecientes a las distintas bases de cribado ya mencionadas fueron protonados a pH 7,4 y posteriormente minimizados en serie mediante dos algoritmos de minimización: *steepest descent*, utilizando los valores por defecto de los parámetros, seguido del gradiente conjugado (en inglés *conjugate gradient*), también con el valor por defecto de sus parámetros. Esta primera etapa preparativa se llevó a cabo con el software OpenBabel versión 2.3 (O'Boyle, 2011).

Seguidamente, los ligandos minimizados fueron convertidos a formato `pdbqt` utilizando el script `prepare_ligand4.py` del paquete MGLTools versión 1.5.7 (Sanner et al., 1996). En el caso de los experimentos con **AD Hyd**, se utilizó el paquete de Python Meeko (Holcomb et al., 2022) para hidratar a los ligandos.

3.8.2.3 PREPARACIÓN DE LOS RECEPTORES

En cuanto a los receptores, la estructura cristalina de **EgFABP1 (PDB: 1O8V)**, como así también las conformaciones de las distintas isoformas obtenidas por **DM** y las estructuras correspondientes a la validación de *docking power* fueron protonadas a

pH 7,4 con el software OpenBabel versión 2.3 (O'Boyle, 2011), y seguidamente convertidas a formato pdbqt utilizando el script *prepare_receptor4.py* del paquete MGLTools versión 1.5.7 (Sanner et al., 1996).

Para los experimentos de *docking power*, las 33 estructuras de las **FABPs** se alinearon estructuralmente usando a **EgFABP1** como referencia, utilizando una grilla cúbica de 22,5 Å (angstroms) de lado con centro en (6,00;9,00;1,00). El espaciado de la grilla (*grid spacing*) fue de 0,375 Å.

Para los experimentos de *scoring power* y **CV**, tanto **EgFABP1** (usada como referencia) como las distintas conformaciones de **DM** fueron alineadas estructuralmente, permitiendo utilizar una grilla cúbica de 22,5 Å de lado, nuevamente, pero con centro en (2,57;-2,66;9,36), utilizando el mismo valor de espaciado en la grilla que en los experimentos de *docking power*. En el caso de los experimentos con **AD Bias**, los **WS** calculados por el algoritmo WATCLUST (López et al., 2015), se introdujeron como diferentes combinaciones de sesgos donores y aceptores de hidrógeno usando el script *prepare_bias.py*, del paquete MGLTools versión 1.5.7 (Sanner et al., 1996). Se utilizó un sesgo de -2 kcal/mol en todos los casos y un radio de decaimiento de 0,9 Å. En todos los experimentos de *docking* se realizaron 20 corridas independientes por ligando, considerando en todos los casos la conformación teórica mejor rankeada, es decir aquella cuyo valor de *score* es máximo en valor absoluto.

3.8.3 VALIDACIÓN DE LOS MODELOS DE DOCKING

3.8.3.1 VALIDACIÓN DE LA FUNCIÓN DE PUNTUACIÓN

Para evaluar la función de puntuación de cada uno de los programas de *docking* utilizados se tomó en cuenta el **AUCROC** (Triballeau et al., 2005) tanto para el conjunto de datos de **LUDe** como de **DUDE-Z**. El programa de *docking* que cuente con el mejor desempeño en ambos conjuntos de datos será el seleccionado para llevar adelante el **CV**. Los **AUCROC** se calcularon utilizando el módulo *roc_auc_score* de la biblioteca *scikit-learn* de Python (Pedregosa et al., 2011).

3.8.3.2 SELECCIÓN DEL PUNTO DE CORTE PARA EL CRIBADO

Una vez seleccionado el algoritmo de *docking* para llevar adelante el **CV** prospectivo, se graficaron las curvas de **PPV** tanto para los conjuntos de **LUDe** y **DUDE-Z**, de manera análoga a los métodos indirectos. De igual manera, el análisis visual de las superficies de **PPV** permitió seleccionar un valor como punto de corte

priorizando la **Sp** por sobre la **Se**, es decir, buscando una baja tasa de falsos positivos.

3.8.3.3 ANÁLISIS DE LOS SITIOS DE UNIÓN PARA CADA UNA DE LAS ISOFORMAS

Luego de implementar el **CV** con las diferentes isoformas de las **FABPs**, se procedió a estudiar las diferencias entre sus respectivos sitios de unión. Para tal fin, se utilizó el servidor DogSiteScorer (Volkamer et al., 2010; Volkamer et al., 2012) para la predicción de cada una de las cavidades y su representación en descriptores moleculares. Muy brevemente, la herramienta DogSiteScorer permite detectar bolsillos de forma automatizada, calculando descriptores vinculados a su tamaño, forma y características químicas. El algoritmo está basado en métodos geométricos, energéticos y de análisis de imágenes. Para los tres métodos se crea una cuadrícula que contiene a la proteína, fijando un área óptima de acuerdo al tamaño de la misma. Luego el algoritmo verifica en la grilla los espacios que quedan “ocupados” y “libres”. Se considera “ocupado” si el punto se encuentra dentro del radio de van der Waals de un átomo de la proteína, de esta forma los bolsillos se identifican como la fusión de los espacios “vacíos” más cercanos. Aquellos con volumen inferior a 100 \AA^3 son descartados.

Los resultados se expresan en términos de volumen, concavidad, superficie, hidrofobicidad, hidrofiliidad, cantidad de átomos donores y aceptores de hidrógeno, y composición aminoacídica del bolsillo, entre otros. Estos descriptores calculados por DogSiteScorer para cada pose de cada isoforma de las **FABPs** fueron primero filtrados por varianza (aquellos cuya varianza era menor al 5% fueron descartados) y por correlación (aquellos descriptores cuya correlación fuera superior a 0,7 también se descartaron). Finalmente, con los descriptores restantes se llevó adelante un análisis de *clustering* vía **PCA** para agrupar aquellos sitios de unión más similares entre sí.

3.8.3.4 VALIDACIÓN DE LA PREDICCIÓN DE LA POSE EXPERIMENTAL

Como se mencionó en la sección 3.8.2.1, la validación del *docking power* se llevó a cabo utilizando 33 complejos **FABP**-ligando diferentes, en donde se calculó el **RMSD** para cada una de las corridas, para cada uno de los programas de *docking* descritos con anterioridad. Los resultados se informan como el promedio de **RMSD** con su desvío estándar asociado, para cada uno de los algoritmos utilizados y se visualizan mediante mapas de calor (*heatmaps* en inglés), los cuales fueron calculados utilizando la biblioteca Seaborn de Python (Waskom, 2021).

3.8.4 CRIBADO VIRTUAL

Después de la validación *in silico* de los modelos, se llevó adelante el CV prospectivo sobre las mismas bases de datos descritas anteriormente en la sección 3.6.

3.9 ENSAYOS EXPERIMENTALES

En este último bloque del presente capítulo se describe la metodología experimental llevada a cabo para la confirmación *in vitro* de los compuestos *hits* obtenidos tanto de la campaña de **CV** desde el ligando como desde la estructura.

Para evaluar los distintos compuestos se emplearon técnicas espectroscópicas que permitieran evidenciar la unión proteína-ligando. Particularmente se realizaron experimentos de desplazamiento monitoreados por fluorescencia como así también de dicroísmo circular (**CD**, del inglés *circular dichroism*) utilizando las isoformas **EgFABP1**, **EmFABP1** y **EmFABP3**. Dichas isoformas se encuentran disponibles en nuestro laboratorio, son estables conformacionalmente *in vitro* y no forman dímeros en solución (Bélgamo et al., 2020).

El protocolo experimental utilizado para los ensayos de fluorescencia fue adaptado del trabajo de Bélgamo et al. (2020), mientras que los experimentos de **CD** fueron propuestos en función de las propiedades espectroscópicas de los compuestos a ensayar.

3.9.1 EXPRESIÓN Y PURIFICACIÓN DE LAS FABPs RECOMBINANTES

Tanto **EmFABP1** como **EmFABP3** fueron sobreexpresadas en la cepa de expresión *Escherichia coli* BL21(DE3) (Novagen, 69450), cuyos *stocks* correspondientes contienen la secuencia codificante (**CDS**) de cada proteína insertada en el sitio de clonado múltiple del vector de expresión pET28a(+) (Novagen, 69864). En el caso de **EgFABP1**, su **CDS** se encuentra en el vector de expresión pET11b (Novagen, 69437). Las **CDS** de cada una de las mencionadas isoformas ya se encontraban disponibles en el laboratorio, como consecuencia de los trabajos previos de Pórfido et al. (2012), y Bélgamo et al. (2020). El agente inductor de la expresión utilizado fue el isopropil- β -D-1-tiogalactopiranosido (**IPTG**, GenBioTech, I2481C5). La composición de las soluciones utilizadas en la purificación, como así también el mapa correspondiente a los vectores pET28a(+) y pET11b se detallan en el **Material Suplementario 3 (Tabla S3d y Figuras S3a-b)**. A continuación se describe el proceso de sobreexpresión y purificación de cada una de las isoformas, mientras que la **Figura 3.9.1** esquematiza todo el proceso.

El protocolo de sobreexpresión de cada una de las isoformas de **FABPs** se llevó adelante en dos días consecutivos. En el día 1, se preparó 1 L de medio *Luria-Bertani* (**LB**), repartido en 5 erlenmeyers de 2 L cada uno. Cada erlenmeyer fue autoclavado y reservado para el día siguiente. Por otro lado, se prepararon 10 mL de medio **LB** para realizar el pre-cultivo, al cual se le agregó 10 μ L del antibiótico Kanamicina (*Gibco*, 11815-032) (1000X, 50 mg/mL) ó Ampicilina (*Sigma*, A9518) (1000X, 100 mg/mL) para el caso de **EgFABP1**. El pre-cultivo fue inoculado con los correspondientes *stocks* en glicerol de BL21(DE3) correspondientes a **EmFABP1**, **EmFABP3** ó **EgFABP1** y puesto en agitación toda la noche a 37°C y 150 rpm.

Al día siguiente, se le agregan 200 μ L de Kanamicina (1000X, 50 mg/mL) ó Ampicilina (1000X, 100 mg/mL), a cada uno de los erlenmeyers de 2L preparados el día anterior. Seguidamente, se inoculan 2 mL del cultivo crecido el día anterior a cada uno de los erlenmeyer e incubar a 37°C y 150 rpm, midiendo cada una hora la densidad óptica (**DO**) del cultivo a 600 nm. Cuando el cultivo se encuentre en fase exponencial (**DO** \geq 0,5), se le agrega a cada erlenmeyer 200 μ L de solución stock de **IPTG** (1000X, 1M) y se incuba por 3 horas a 37°C y 150 rpm. Finalizada la incubación, se centrifugó el cultivo a 2500 g por 20 minutos y seguidamente se descartó el sobrenadante, mientras que el pellet fue resuspendido en 20 mL de buffer de unión (composición disponible en **Tabla S3d**). Por último, el pellet resuspendido se guardó a -80°C en un tubo Falcon de 50 mL, para la etapa de purificación.

Una vez sobreexpresada cada isoforma, se procedió a la etapa de purificación. Para ello, las células fueron lisadas mediante sonicación y posterior centrifugación a 8000 g durante 40 minutos a 4°C. Dado que **EmFABP1** y **EmFABP3** fueron subclonadas en el vector de expresión pET-28a(+), contienen en su extremo C-terminal una cola de histidinas (*His-Tag*) lo cual permite su purificación mediante una columna de afinidad, compuesta por una matriz de níquel. Por otra parte, **EgFABP1** no cuenta con *His-Tag*, razón por la cual este primer paso de purificación no fue realizado para esta isoforma. Luego, se realizó una cromatografía de exclusión molecular utilizando una columna Superdex-75 (*GE Healthcare*, 17-5174-01) y finalmente, un último paso de deslipidación utilizando una cromatografía hidrofóbica Lipidex-100 (*Sigma*, H6383). En el caso de **EgFABP1**, la columna de exclusión molecular Sephadex-G50 (*Pharmacia Biotech Inc.*, 17-0042-02) (Pórfido et al., 2012) fue la utilizada en lugar de la Superdex-75. La pureza de la proteína a lo largo de la marcha de purificación fue evaluada mediante geles de poliacrilamida al 15% con dodecil sulfato de sodio (**SDS-PAGE** 15%).

El protocolo de purificación empieza con el descongelamiento del pellet bacteriano, el cual es retirado del ultrafreezer y guardado a -20°C la noche anterior a la lisis celular. A la mañana siguiente, la muestra es sonicada utilizando un sonicador de punta (Branson 450 CE, *Wolf labs*, 159-063-676) durante 3 ciclos de 5 minutos al 50% de la potencia en baño de hielo. Se reservan 50 μ L del homogenato para

sembrar en gel de **SDS-PAGE** 15%, y el resto de la muestra se centrifuga a 8000 g por 40 minutos a 4°C. Seguidamente, se separa el pellet del sobrenadante, reservando 50 µL de cada fracción para su posterior chequeo por **SDS-PAGE** 15%. Para el caso de **EmFABP1** y **EmFABP3**, se continuó con la fracción sobrenadante hacia la columna de afinidad. Para **EgFABP1** se llevó a cabo directamente la cromatografía de exclusión molecular con su fracción sobrenadante, dada la ausencia de *His-Tag*.

La columna de afinidad se lavó primeramente con agua (10 volúmenes de columna, **VC**) y luego con buffer de unión (10 **VC**). Seguidamente, se adiciona todo el sobrenadante a través de la columna de afinidad, colectando el flujo que sale de la columna (fracción no retenida, **NR**). A continuación, se lava con 10 **VC** de buffer de unión y se recolecta en la misma fracción **NR**. Posteriormente, se trata a la columna con 6 **VC** de buffer de lavado, colectando en una nueva fracción (fracción lavado, **W**). Por último, se realizan 6 lavados de 4 mL cada uno con buffer de elución, recolectando cada uno de ellos en una fracción diferente (fracciones **E1-E6**). Se toman 10 µL de cada una de las fracciones más 2 µL de buffer de carga 5X y se siembran dichas fracciones en un gel **SDS PAGE** 15%. La electroforesis se corre durante 1,5 horas a 150 V.

Al día siguiente, se colectan las fracciones en donde se encuentre la proteína, de acuerdo al gel de **SDS-PAGE** del día anterior y se concentra la muestra a 2 mL totales, utilizando un filtro de centrifuga de 3 kDa de tamaño de poro (*Merck Millipore*, UFC9003). Se inyectan 500 µL de muestra en la columna Superdex-75, colectando fracciones de 0,7 mL cada una. Repetir el proceso hasta pasar toda la muestra. En el caso de **EgFABP1**, sembrar la totalidad de la muestra en la columna G-50, mediante una bomba peristáltica P-1 (*Pharmacia Biotech Inc.*), a una velocidad de flujo de 1,5 mL/minuto, colectando fracciones de 10 mL cada una. Sembrar las fracciones candidatas a tener proteína (aumento en la absorbancia a 280 nm del detector del equipo) en un gel de **SDS-PAGE** 15%, y correr la electroforesis en las mismas condiciones anteriormente mencionadas. Guardar las fracciones obtenidas de la cromatografía en la heladera.

Al otro día, se colectan las fracciones con proteína, de acuerdo a los resultados del gel anterior y concentrar la muestra a 6 mL totales, de la misma manera que se mencionó anteriormente. La muestra se pasa por la columna Lipidex-100, a un flujo de 1mL/min a 37°C, colectando fracciones cada 3 minutos, obteniendo 20 fracciones en total y midiendo la absorbancia a 280 nm para cada una de ellas. Se siembran en un gel de **SDS-PAGE** 15% las fracciones con absorbancia apreciable y se corre la electroforesis. Finalmente, se colectan las fracciones que contienen la proteína de interés, se concentra a 5 mL y se alicuota la muestra en tubos eppendorf de 1,5 mL, para finalmente almacenar la proteína purificada a -80°C.

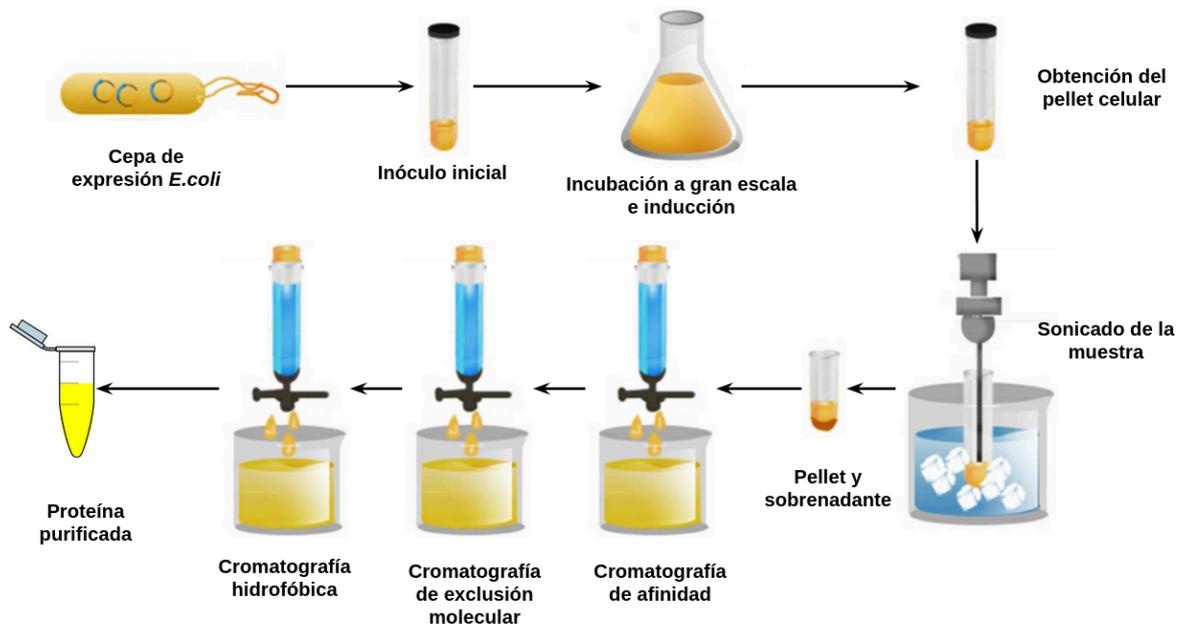


Figura 3.9.1. Esquema simplificado del protocolo de purificación de proteínas seguido en esta tesis doctoral. Para el caso de **EgFABP1** no se realiza cromatografía de afinidad.

3.9.2 ENSAYOS DE DESPLAZAMIENTO MONITOREADOS POR FLUORESCENCIA

La fluorescencia es un fenómeno físico en el cual ciertas sustancias, llamadas fluoróforos, absorben energía de una fuente externa, y emiten luz a una longitud de onda mayor. Esta emisión de luz se produce debido a la relajación de los electrones excitados a su estado fundamental. Este fenómeno se utiliza para estudiar interacciones proteína-ligando, visualizar procesos celulares, analizar expresión génica, detectar y cuantificar biomarcadores, entre muchas otras aplicaciones (Lakowicz, 2006).

En el contexto de la presente tesis doctoral, los *hits* obtenidos *in silico* tanto por métodos indirectos y directos fueron evaluados mediante la técnica de desplazamiento por fluorescencia, utilizando un protocolo adaptado del trabajo de Bélgamo et al. (2020).

Los ensayos de desplazamiento monitoreados por fluorescencia se basan en utilizar sondas fluorescentes, las cuales tienen la propiedad de emitir fluorescencia únicamente cuando se encuentran unidas al sitio activo de interés. Consecuentemente, la fluorescencia generada por estas sondas puede ser “desplazada” o modificada debido a la interacción de la proteína con otro ligando competidor. La adición de ligando competidor a diferentes concentraciones y el posterior cambio en la fluorescencia de la sonda permite la medición de constantes de disociación aparentes (K_{dapp}) para cada uno de los ligandos ensayados,

permitiendo estimar la potencia de cada uno de los compuestos frente al blanco de interés.

Para nuestros propósitos, la sonda elegida para los experimentos de desplazamiento fue el **ANS** (*Invitrogen*, J65538.22). Como punto de partida del experimento, las condiciones utilizadas en cubeta fueron: **ANS** 20 μM en presencia de 1 μM de proteína en buffer fosfatos (**PBS**) a $\text{pH}=7,4$, en un volumen final de 200 μL . A este complejo inicial se le realizaron agregados sucesivos de los distintos ligandos a ensayar hasta no observar diferencias en la fluorescencia emitida. Para cada uno de los puntos se utilizó una longitud de onda de excitación de 350 nm, mientras que se colectó el espectro de emisión entre 380-600 nm. Como control positivo de desplazamiento, se utilizó al Ácido oleico, un ligando conocido de las **FABPs** (Richieri et al., 1994; Zimmerman & Veerkamp, 2002; Bélgamo et al., 2020). Entre agregados, el tiempo de incubación fue de 2 min y la temperatura 25°C. Un esquema del procedimiento se muestra en la **Figura 3.9.2**.

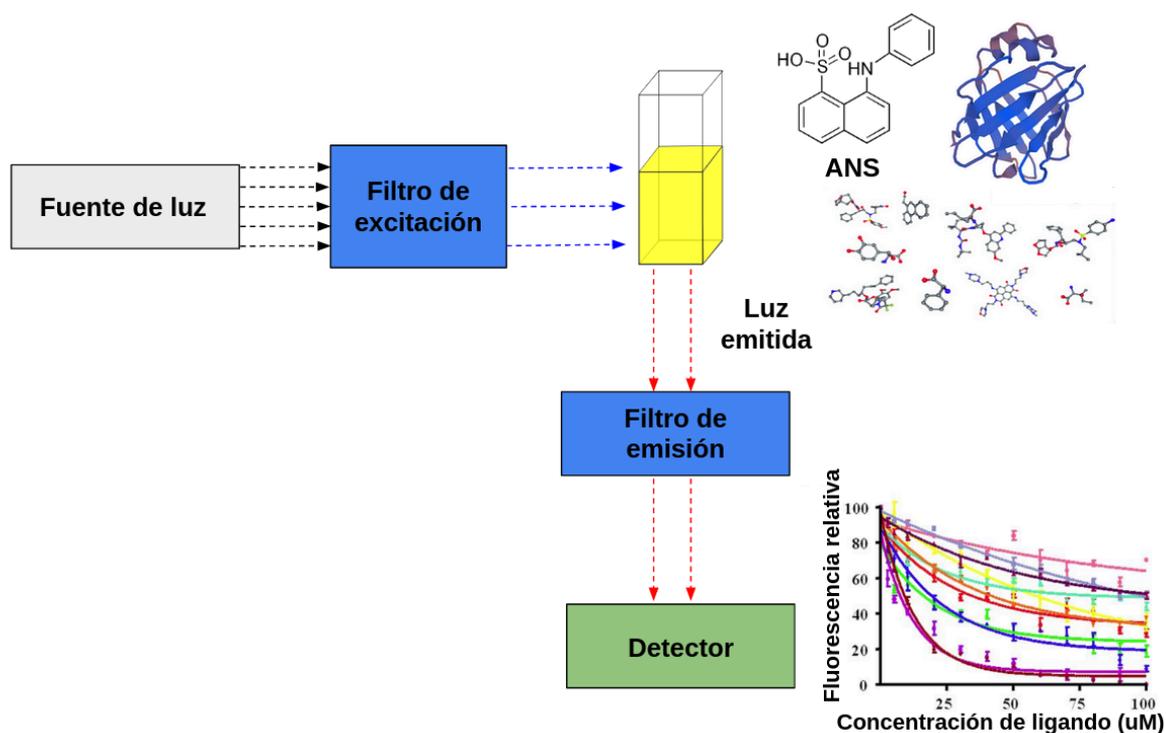
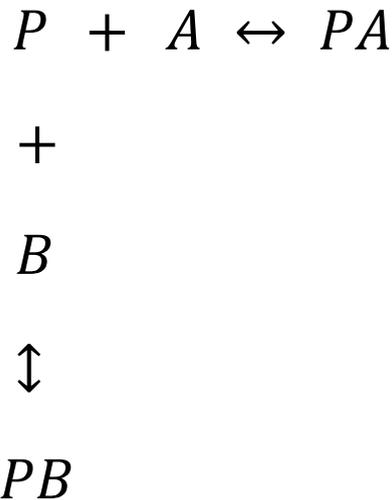


Figura 3.9.2. Esquema simplificado de los ensayos de desplazamiento de fluorescencia. El gráfico final de fluorescencia relativa en función de la concentración de ligando fue adaptado de Zhang et al. (2014).

3.9.3 AJUSTE ESTADÍSTICO Y DETERMINACIÓN DE CONSTANTES DE DISOCIACIÓN

Los valores de fluorescencia obtenidos para cada par proteína-compuesto en función de la concentración de este último fueron ajustados de manera no lineal

considerando un modelo que tiene en cuenta un solo sitio de unión a ligando, algo ya reportado para algunas de las **FABPs** (Bélgamo et al., 2020; Franchini et al., 2009). De esta manera, se puede plantear un equilibrio entre la proteína (P) y la sonda fluorescente (A) en donde se formará el complejo PA , el cual emitirá fluorescencia, y otro equilibrio en donde el compuesto desplazante (B) formará el complejo PB :



Conforme aumenta la cantidad de B agregada, la cantidad de complejo PA disminuirá, y con ello, la señal de fluorescencia. En consecuencia, cuando la concentración de B_0 (esto es, la suma de B y PB) sea cero, la concentración de PA será máxima, y por lo tanto mostrará una fluorescencia máxima, y por el contrario, a altas concentraciones de B_0 se espera una fluorescencia residual, como se puede apreciar en la **Figura 3.9.3**.

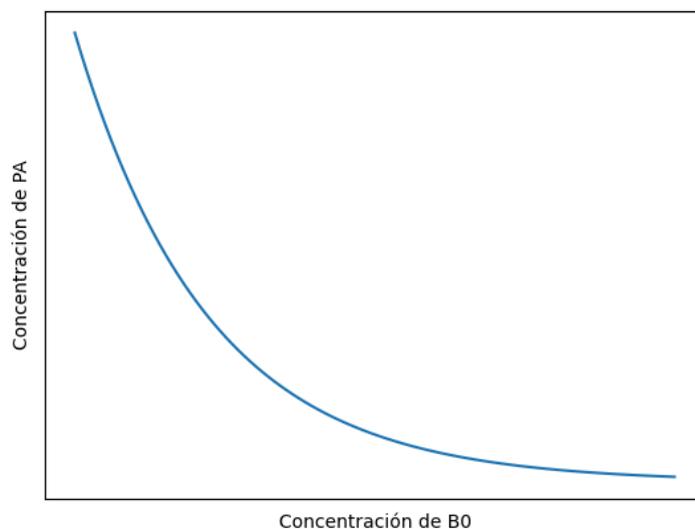


Figura 3.9.3. Gráfico típico de un experimento de desplazamiento por fluorescencia, con un único sitio de unión a ligando.

Partiendo de los equilibrios recién planteados, se puede llegar a la **Ecuación 3.10** (la deducción completa se encuentra en la sección **Formalismo S3 del Material Suplementario 3**):

$$\Delta F = \frac{\Delta F_0}{1 + \frac{B}{Kd_{app}}} + \Delta F_{res} \quad (3.10)$$

donde, ΔF se corresponde a la fluorescencia observada correspondiente a la concentración de ligando desplazante B , ΔF_0 es la diferencia de fluorescencia entre las concentraciones mínima y máxima de B ($B=0$ y $B=B$ máx) y ΔF_{res} es un término que tiene en cuenta la fluorescencia residual a altas concentraciones de ligando. Todos los valores de fluorescencia fueron normalizados por la fluorescencia máxima $\Delta F_{max} = \Delta F_0 + \Delta F_{res}$. Los datos obtenidos para cada una de las curvas proteína-compuesto fueron ajustados a la **Ecuación 3.10**, obteniendo Kd_{app} para cada caso utilizando la herramienta *Solver* en el programa Excel (Gutiérrez Villaverde, 2021).

3.9.4 ENSAYOS DE DICROÍSMO CIRCULAR

El **CD** es una técnica que se basa en la diferencia de absorbancia de la luz circularmente polarizada (izquierda y derecha) por una muestra (Campbell, 2012). Cuando la molécula en cuestión exhibe estructuras asimétricas y/o quiralidad, puede interactuar de manera diferencial con la luz polarizada en sentido horario (derecha) y antihorario (izquierda).

Esta diferencia en la absorción de la luz circularmente polarizada está relacionada con la conformación tridimensional y las interacciones moleculares de la muestra. En el caso de proteínas, el **CD** puede proporcionar información sobre la estructura secundaria (α -hélices, hojas β , loops) y cambios conformacionales inducidos por factores como cambios de pH o unión a ligandos (Prince & Nairn, 2009).

El **CD** se suele describir en dos partes, en función de la región del espectro UV analizado:

1. **CD en el UV lejano (far-UV CD en inglés)**: Se trabaja en el rango del espectro entre 190 y 250 nm. El **far-UV CD** es especialmente útil para analizar la estructura secundaria de proteínas y ácidos nucleicos. Proporciona información sobre la cantidad y el tipo de estructuras secundarias presentes en la molécula, como α -hélices, hojas β y loops. Los cambios en la intensidad y forma de las bandas de absorción en este rango pueden indicar alteraciones en la conformación secundaria debido a cambios en el entorno o a la unión de ligandos (Prince & Nairn, 2009).

2. ***CD en el UV cercano (near-UV CD en inglés)***: En el caso del ***near-UV CD***, se emplea radiación en el rango entre 250 y 350 nm. Esta técnica se utiliza para estudiar las interacciones entre cromóforos aromáticos (por ejemplo, los aminoácidos tirosina y triptófano) y para analizar la estructura terciaria y los cambios conformacionales en las proteínas. La absorción en esta región está influenciada por la disposición tridimensional de los cromóforos aromáticos y sus interacciones con el entorno. Cambios en el entorno local de estos aminoácidos, como la unión a ligandos o alteraciones conformacionales, pueden ser detectados y analizados (Prince & Nairn, 2009).

En este trabajo de tesis doctoral, se enfrentaron las isoformas **EgFABP1**, **EmFABP1** y **EmFABP3** en distintas relaciones molares con respecto a los diferentes compuestos a ensayar en el ***far-UV CD***. Los espectros fueron tomados utilizando una cubeta de 0,1 cm de camino óptico, a una velocidad de barrido de 20 nm/min con una constante de tiempo de 1s a 25°C de temperatura. Los espectros obtenidos fueron subsecuentemente normalizados por la concentración de cada proteína y el camino óptico empleado.

REFERENCIAS

Abraham, M. J., Murtola, T., Schulz, R., Páll, S., Smith, J. C., Hess, B., & Lindahl, E. (2015). GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX*, 1–2, 19–25. <https://doi.org/10.1016/j.softx.2015.06.001>

Alberca, Lucas N. (2018). *Búsqueda asistida por computadora de nuevos fármacos antichagásicos análogos de poliaminas* [Doctor en Ciencias Exactas, área Ciencias Biológicas, Universidad Nacional de La Plata]. <https://doi.org/10.35537/10915/72607>

Alberca, L. N., Sbaraglini, M. L., Morales, J. F., Dietrich, R., Ruiz, M. D., Pino Martínez, A. M., Miranda, C. G., Fraccaroli, L., Alba Soto, C. D., Carrillo, C., Palestro, P. H., & Talevi, A. (2018). Cascade Ligand- and Structure-Based Virtual Screening to Identify New Trypanocidal Compounds Inhibiting Putrescine Uptake. *Frontiers in Cellular and Infection Microbiology*, 8, 173. <https://doi.org/10.3389/fcimb.2018.00173>

Amaro, R. E., Baudry, J., Chodera, J., Demir, Ö., McCammon, J. A., Miao, Y., & Smith, J. C. (2018). Ensemble Docking in Drug Discovery. *Biophysical Journal*, 114(10), 2271–2278. <https://doi.org/10.1016/j.bpj.2018.02.038>

Arcon, J. P., Modenutti, C. P., Avendaño, D., Lopez, E. D., Defelipe, L. A., Ambrosio, F. A., Turjanski, A. G., Forli, S., & Marti, M. A. (2019). AutoDock Bias: Improving binding mode prediction and virtual screening using known protein–ligand interactions. *Bioinformatics*, 35(19), 3836–3838. <https://doi.org/10.1093/bioinformatics/btz152>

Bélgamo, J. A., Alberca, L. N., Pórfido, J. L., Romero, F. N. C., Rodriguez, S., Talevi, A., Córscico, B., & Franchini, G. R. (2020). Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets. *Journal of Computer-Aided Molecular Design*, 34(12), 1275–1288. <https://doi.org/10.1007/s10822-020-00352-8>

Bemis, G. W., & Murcko, M. A. (1996). The Properties of Known Drugs. 1. Molecular Frameworks. *Journal of Medicinal Chemistry*, 39(15), 2887–2893. <https://doi.org/10.1021/jm9602928>

Berendsen, H. J. C., Van Der Spoel, D., & Van Drunen, R. (1995). GROMACS: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91(1–3), 43–56. [https://doi.org/10.1016/0010-4655\(95\)00042-E](https://doi.org/10.1016/0010-4655(95)00042-E)

Böcker, A. (2008). Toward an Improved Clustering of Large Data Sets Using Maximum Common Substructures and Topological Fingerprints. *Journal of Chemical Information and Modeling*, 48(11), 2097–2107. <https://doi.org/10.1021/ci8000887>

Bussi, G., Donadio, D., & Parrinello, M. (2007). Canonical sampling through velocity rescaling. *The Journal of Chemical Physics*, 126(1), 014101. <https://doi.org/10.1063/1.2408420>

Campbell, I. D. (2012). *Biophysical techniques*. Oxford university press.

Carbonneau, M.-A., Granger, E., Raymond, A. J., & Gagnon, G. (2016). Robust multiple-instance learning ensembles using random subspace instance selection. *Pattern Recognition*, 58, 83–99. <https://doi.org/10.1016/j.patcog.2016.03.035>

Chemaxon. (2023). *MarvinJS 17.21.0*. Retrieved September 14, 2023, from <https://chemaxon.com>

Choueiry, G. (2022). Understand Forward and Backward Stepwise Regression. *Quantifying Health*. <https://quantifyinghealth.com/stepwise-selection/>

Computational Chemistry Group, CEQUINOR, UNLP. (2023). *NaturAr*. Retrieved September 14, 2023, from <https://naturar.quimica.unlp.edu.ar/en/about/>

Corsello, S. M., Bittker, J. A., Liu, Z., Gould, J., McCarren, P., Hirschman, J. E., Johnston, S. E., Vrcic, A., Wong, B., Khan, M., Asiedu, J., Narayan, R., Mader, C. C., Subramanian, A., & Golub, T. R. (2017). The Drug Repurposing Hub: A next-generation drug library and information resource. *Nature Medicine*, 23(4), 405–408. <https://doi.org/10.1038/nm.4306>

Daho, M. E. H., & Chikh, M. A. (2015). Combining Bootstrapping Samples, Random Subspaces and Random Forests to Build Classifiers. *Journal of Medical Imaging and Health Informatics*, 5(3), 539–544. <https://doi.org/10.1166/jmih.2015.1423>

Davies, M., Nowotka, M., Papadatos, G., Dedman, N., Gaulton, A., Atkinson, F., Bellis, L., & Overington, J. P. (2015). ChEMBL web services: Streamlining access to drug discovery data and utilities. *Nucleic Acids Research*, 43(W1), W612–W620. <https://doi.org/10.1093/nar/gkv352>

Davis & Ward. (2014). *The Handbook of Medicinal Chemistry: Principles and Practice*. Royal Society of Chemistry. <https://books.google.com.ar/books?id=mICsEAAAQBAJ>

Dimitrov, S., Dimitrova, G., Pavlov, T., Dimitrova, N., Patlewicz, G., Niemela, J., & Mekenyan, O. (2005). A Stepwise Approach for Defining the Applicability Domain of SAR and QSAR Models. *Journal of Chemical Information and Modeling*, 45(4), 839–849. <https://doi.org/10.1021/ci0500381>

Eberhardt, J., Santos-Martins, D., Tillack, A. F., & Forli, S. (2021). AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *Journal*

Hawkins, D. M., Kraker, J. J., Basak, S. C., & Mills, D. (2008). QSPR checking and validation: A case study with hydroxy radical reaction rate constant. *SAR and QSAR in Environmental Research*, 19(5–6), 525–539. <https://doi.org/10.1080/10629360802349058>

Hess, B., Kutzner, C., Van Der Spoel, D., & Lindahl, E. (2008). GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *Journal of Chemical Theory and Computation*, 4(3), 435–447. <https://doi.org/10.1021/ct700301q>

Holcomb, M., Santos-Martins, D., Tillack, A. F., & Forli, S. (2022). Performance evaluation of flexible macrocycle docking in AutoDock. *QRB Discovery*, 3, e18. <https://doi.org/10.1017/qrd.2022.18>

Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1), 33–38. [https://doi.org/10.1016/0263-7855\(96\)00018-5](https://doi.org/10.1016/0263-7855(96)00018-5)

Izadi, S., Anandakrishnan, R., & Onufriev, A. V. (2014, October 23). Building Water Models: A Different Approach. *The Journal of Physical Chemistry Letters*, 5(21), 3863–3871. <https://doi.org/10.1021/jz501780a>

Jakobsson, E., Alvite, G., Bergfors, T., Esteves, A., & Kleywegt, G. J. (2003). The crystal structure of Echinococcus granulosus fatty-acid-binding protein 1. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1649(1), 40–50. [https://doi.org/10.1016/S1570-9639\(03\)00151-1](https://doi.org/10.1016/S1570-9639(03)00151-1)

Jarvis, R. A., & Patrick, E. A. (1973). Clustering Using a Similarity Measure Based on Shared Near Neighbors. *IEEE Transactions on Computers*, C-22(11), 1025–1034. <https://doi.org/10.1109/T-C.1973.223640>

Jaworska, J., Nikolova-Jeliazkova, N., & Aldenberg, T. (2005). QSAR Applicability Domain Estimation by Projection of the Training Set in Descriptor Space: A Review. *Alternatives to Laboratory Animals*, 33(5), 445–459. <https://doi.org/10.1177/026119290503300508>

Jorgensen, W. L., & Tirado-Rives, J. (1988). The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. *Journal of the American Chemical Society*, 110(6), 1657–1666. <https://doi.org/10.1021/ja00214a001>

Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., Tunyasuvunakool, K., Bates, R., Žídek, A., Potapenko, A., Bridgland, A., Meyer, C., Kohl, S. A. A., Ballard, A. J., Cowie, A., Romera-Paredes, B., Nikolov, S., Jain, R., Adler, J., ... Hassabis, D. (2021). Highly accurate protein structure prediction with

AlphaFold. *Nature*, 596(7873), 583–589.
<https://doi.org/10.1038/s41586-021-03819-2>

Kirchmair, J., Markt, P., Distinto, S., Wolber, G., & Langer, T. (2008). Evaluation of the performance of 3D virtual screening protocols: RMSD comparisons, enrichment assessments, and decoy selection—What can we learn from earlier mistakes? *Journal of Computer-Aided Molecular Design*, 22(3–4), 213–228.
<https://doi.org/10.1007/s10822-007-9163-6>

Klug, D. M., Gelb, M. H., & Pollastri, M. P. (2016). Repurposing strategies for tropical disease drug discovery. *Bioorganic & Medicinal Chemistry Letters*, 26(11), 2569–2576. <https://doi.org/10.1016/j.bmcl.2016.03.103>

Lakowicz, J. R. (2010). *Principles of fluorescence spectroscopy* (Third edition, corrected at 4. printing). Springer.

Lee, C.-H., Lui, D. T. W., & Lam, K. S. L. (2021). Adipocyte Fatty Acid-Binding Protein, Cardiovascular Diseases and Mortality. *Frontiers in Immunology*, 12, 589206. <https://doi.org/10.3389/fimmu.2021.589206>

Li, L., Hu, Q., Wu, X., & Yu, D. (2014). Exploration of classification confidence in ensemble learning. *Pattern Recognition*, 47(9), 3120–3131.
<https://doi.org/10.1016/j.patcog.2014.03.021>

Li, Y., Zhao, C., Zhang, J., Zhai, S., Wei, B., & Wang, L. (2019). HybridMolDB: A Manually Curated Database Dedicated to Hybrid Molecules for Chemical Biology and Drug Discovery. *Journal of Chemical Information and Modeling*, 59(10), 4063–4069. <https://doi.org/10.1021/acs.jcim.9b00314>

Lindahl, E., Hess, B., & Van Der Spoel, D. (2001). GROMACS 3.0: A package for molecular simulation and trajectory analysis. *Journal of Molecular Modeling*, 7(8), 306–317. <https://doi.org/10.1007/s008940100045>

Llanos, M. A., Gantner, M. E., Rodriguez, S., Alberca, L. N., Bellera, C. L., Talevi, A., & Gavernet, L. (2021). Strengths and Weaknesses of Docking Simulations in the SARS-CoV-2 Era: The Main Protease (Mpro) Case Study. *Journal of Chemical Information and Modeling*, 61(8), 3758–3770.
<https://doi.org/10.1021/acs.jcim.1c00404>

Llanos, M. A., Sbaraglini, M. L., Villalba, M. L., Ruiz, M. D., Carrillo, C., Alba Soto, C., Talevi, A., Angeli, A., Parkkila, S., Supuran, C. T., & Gavernet, L. (2020). A structure-based approach towards the identification of novel antichagasic compounds: *Trypanosoma cruzi* carbonic anhydrase inhibitors. *Journal of Enzyme Inhibition and Medicinal Chemistry*, 35(1), 21–30.
<https://doi.org/10.1080/14756366.2019.1677638>

López, E. D., Arcon, J. P., Gauto, D. F., Petruk, A. A., Modenutti, C. P., Dumas, V. G., Marti, M. A., & Turjanski, A. G. (2015). WATCLUST: A tool for improving the design of drugs based on protein-water interactions. *Bioinformatics*, *31*(22), 3697–3699. <https://doi.org/10.1093/bioinformatics/btv411>

Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics (Oxford, England)*, *29*(21), 2722–2728. <https://doi.org/10.1093/bioinformatics/btt473>

Martin, T. M., Harten, P., Young, D. M., Muratov, E. N., Golbraikh, A., Zhu, H., & Tropsha, A. (2012). Does Rational Selection of Training and Test Sets Improve the Outcome of QSAR Modeling? *Journal of Chemical Information and Modeling*, *52*(10), 2570–2578. <https://doi.org/10.1021/ci300338w>

Min, S.-H. (2016). *A Genetic Algorithm-Based Heterogeneous Random Subspace Ensemble Model for Bankruptcy Prediction*. <https://www.semanticscholar.org/paper/A-Genetic-Algorithm-Based-Heterogeneous-Random-for-Min/4dbaa575bae10e1de60d047ed9006b32a550fa82#citing-papers>

Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., & Steinegger, M. (2022). ColabFold: Making protein folding accessible to all. *Nature Methods*, *19*(6), 679–682. <https://doi.org/10.1038/s41592-022-01488-1>

Morales, J. F., Chuguransky, S., Alberca, L. N., Alice, J. I., Goicoechea, S., Ruiz, M. E., Bellera, C. L., & Talevi, A. (2020). Positive Predictive Value Surfaces as a Complementary Tool to Assess the Performance of Virtual Screening Methods. *Mini-Reviews in Medicinal Chemistry*, *20*(14), 1447–1460. <https://doi.org/10.2174/1871525718666200219130229>

Moriwaki, H., Tian, Y.-S., Kawashita, N., & Takagi, T. (2018). Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, *10*(1), 4. <https://doi.org/10.1186/s13321-018-0258-y>

National Institutes of Health (NIH). (2023). *PubMed*. PubMed. Retrieved September 14, 2023, from <https://pubmed.ncbi.nlm.nih.gov/>

Novick, P. A., Ortiz, O. F., Poelman, J., Abdulhay, A. Y., & Pande, V. S. (2013). SWEETLEAD: An In Silico Database of Approved Drugs, Regulated Chemicals, and Herbal Isolates for Computer-Aided Drug Discovery. *PLoS ONE*, *8*(11), e79568. <https://doi.org/10.1371/journal.pone.0079568>

O'Boyle, N. M., Banck, M., James, C. A., Morley, C., Vandermeersch, T., & Hutchison, G. R. (2011). Open Babel: An open chemical toolbox. *Journal of Cheminformatics*, *3*(1), 33. <https://doi.org/10.1186/1758-2946-3-33>

Páll, S., Abraham, M. J., Kutzner, C., Hess, B., & Lindahl, E. (2015). Tackling Exascale Software Challenges in Molecular Dynamics Simulations with GROMACS. In S. Markidis & E. Laure (Eds.), *Solving Software Challenges for Exascale* (Vol. 8759, pp. 3–27). Springer International Publishing. https://doi.org/10.1007/978-3-319-15976-8_1

Palma, J., & Pierdominici-Sottile, G. (2023). On the Uses of PCA to Characterise Molecular Dynamics Simulations of Biological Macromolecules: Basics and Tips for an Effective Use. *ChemPhysChem*, 24(2), e202200491. <https://doi.org/10.1002/cphc.202200491>

Parrinello, M., & Rahman, A. (1981). Polymorphic transitions in single crystals: A new molecular dynamics method. *Journal of Applied Physics*, 52(12), 7182–7190. <https://doi.org/10.1063/1.328693>

Pearce, R., & Zhang, Y. (2021). Toward the solution of the protein structure prediction problem. *Journal of Biological Chemistry*, 297(1), 100870. <https://doi.org/10.1016/j.jbc.2021.100870>

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., & Cournapeau, D. (2011). Scikit-learn: Machine Learning in Python. *MACHINE LEARNING IN PYTHON*.

Peduzzi, P., Concato, J., Feinstein, A. R., & Holford, T. R. (1995). Importance of events per independent variable in proportional hazards regression analysis II. Accuracy and precision of regression estimates. *Journal of Clinical Epidemiology*, 48(12), 1503–1510. [https://doi.org/10.1016/0895-4356\(95\)00048-8](https://doi.org/10.1016/0895-4356(95)00048-8)

Peduzzi, P., Concato, J., Kemper, E., Holford, T. R., & Feinstein, A. R. (1996). A simulation study of the number of events per variable in logistic regression analysis. *Journal of Clinical Epidemiology*, 49(12), 1373–1379. [https://doi.org/10.1016/S0895-4356\(96\)00236-3](https://doi.org/10.1016/S0895-4356(96)00236-3)

Perutz, M. F., Rossmann, M. G., Cullis, A. F., Muirhead, H., Will, G., & North, A. C. T. (1960). Structure of Hæmoglobin: A Three-Dimensional Fourier Synthesis at 5.5-Å. Resolution, Obtained by X-Ray Analysis. *Nature*, 185(4711), 416–422. <https://doi.org/10.1038/185416a0>

Pilon, A. C., Valli, M., Dametto, A. C., Pinto, M. E. F., Freire, R. T., Castro-Gamboa, I., Andricopulo, A. D., & Bolzani, V. S. (2017). NuBBEDB: An updated database to uncover chemical and biological information from Brazilian biodiversity. *Scientific Reports*, 7(1), 7215. <https://doi.org/10.1038/s41598-017-07451-x>

Porfido, J. L., Alvite, G., Silva, V., Kennedy, M. W., Esteves, A., & Corsico, B. (2012). Direct Interaction between EgFABP1, a Fatty Acid Binding Protein from

Echinococcus granulosus, and Phospholipid Membranes. *PLoS Neglected Tropical Diseases*, 6(11), e1893. <https://doi.org/10.1371/journal.pntd.0001893>

Pórfido, J. L., Herz, M., Kiss, F., Kamenetzky, L., Brehm, K., Rosenzvit, M. C., Córscico, B., & Franchini, G. R. (2020). Fatty acid-binding proteins in Echinococcus spp.: The family has grown. *Parasitology Research*, 119(4), 1401–1408. <https://doi.org/10.1007/s00436-020-06631-5>

Prada Gori, D. N., Alberca, L. N., Rodriguez, S., Alice, J. I., Llanos, M. A., Bellera, C. L., & Talevi, A. (2022b). LIDeB Tools: A Latin American resource of freely available, open-source cheminformatics apps. *Artificial Intelligence in the Life Sciences*, 2, 100049. <https://doi.org/10.1016/j.aillsci.2022.100049>

Prada Gori, D. N., Llanos, M. A., Bellera, C. L., Talevi, A., & Alberca, L. N. (2022a). iRaPCA and SOMoC: Development and Validation of Web Applications for New Approaches for the Clustering of Small Molecules. *Journal of Chemical Information and Modeling*, 62(12), 2987–2998. <https://doi.org/10.1021/acs.jcim.2c00265>

Price, N. C., & Nairn, J. (2009). *Exploring proteins: A student's guide to experimental skills and methods*. Oxford Univ. Press.

Pronk, S., Páll, S., Schulz, R., Larsson, P., Bjelkmar, P., Apostolov, R., Shirts, M. R., Smith, J. C., Kasson, P. M., Van Der Spoel, D., Hess, B., & Lindahl, E. (2013). GROMACS 4.5: A high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, 29(7), 845–854. <https://doi.org/10.1093/bioinformatics/btt055>

Quiroga, R., & Villarreal, M. A. (2016). Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLOS ONE*, 11(5), e0155183. <https://doi.org/10.1371/journal.pone.0155183>

Rational Discovery LLC. (2006). *RDKit: Open-Source Cheminformatics and Machine Learning Software, Open-Source Cheminformatics and Machine Learning*. <http://www.rdkit.org/>

Richieri, G. V., Ogata, R. T., & Kleinfeld, A. M. (1994). Equilibrium constants for the binding of fatty acids with fatty acid-binding proteins from adipocyte, intestine, heart, and liver measured with the fluorescent probe ADIFAB. *The Journal of Biological Chemistry*, 269(39), 23918–23930.

Rogers, D., & Hahn, M. (2010). Extended-Connectivity Fingerprints. *Journal of Chemical Information and Modeling*, 50(5), 742–754. <https://doi.org/10.1021/ci100050t>

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53–65. [https://doi.org/10.1016/0377-0427\(87\)90125-7](https://doi.org/10.1016/0377-0427(87)90125-7)

Rücker, C., Rücker, G., & Meringer, M. (2007). Y-Randomization and Its Variants in QSPR/QSAR. *Journal of Chemical Information and Modeling*, 47(6), 2345–2357. <https://doi.org/10.1021/ci700157b>

Sahigara, F., Mansouri, K., Ballabio, D., Mauri, A., Consonni, V., & Todeschini, R. (2012). Comparison of Different Approaches to Define the Applicability Domain of QSAR Models. *Molecules*, 17(5), 4791–4810. <https://doi.org/10.3390/molecules17054791>

Sanner, M. F., Olson, A. J., & Spehner, J. C. (1996). Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers*, 38(3), 305–320. [https://doi.org/10.1002/\(SICI\)1097-0282\(199603\)38:3%3C305::AID-BIP4%3E3.0.CO;2-Y](https://doi.org/10.1002/(SICI)1097-0282(199603)38:3%3C305::AID-BIP4%3E3.0.CO;2-Y)

Santos-Martins, D., Solis-Vasquez, L., Tillack, A. F., Sanner, M. F., Koch, A., & Forli, S. (2021). Accelerating AutoDock 4 with GPUs and Gradient-Based Local Search. *Journal of Chemical Theory and Computation*, 17(2), 1060–1073. <https://doi.org/10.1021/acs.jctc.0c01006>

Smathers, R. L., & Petersen, D. R. (2011). The human fatty acid-binding protein family: Evolutionary divergences and functions. *Human Genomics*, 5(3), 170. <https://doi.org/10.1186/1479-7364-5-3-170>

Sorokina, M., Merseburger, P., Rajan, K., Yirik, M. A., & Steinbeck, C. (2021). COCONUT online: Collection of Open Natural Products database. *Journal of Cheminformatics*, 13(1), 2. <https://doi.org/10.1186/s13321-020-00478-9>

Stein, R. M., Yang, Y., Balius, T. E., O'Meara, M. J., Lyu, J., Young, J., Tang, K., Shoichet, B. K., & Irwin, J. J. (2021). Property-Unmatched Decoys in Docking Benchmarks. *Journal of Chemical Information and Modeling*, 61(2), 699–714. <https://doi.org/10.1021/acs.jcim.0c00598>

Sterling, T., & Irwin, J. J. (2015). ZINC 15 – Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling*, 55(11), 2324–2337. <https://doi.org/10.1021/acs.jcim.5b00559>

Swain, M. (2019). *MolVS: Molecule Validation and Standardization—MolVS 0.1.1 documentation*. <https://molvs.readthedocs.io/en/latest/>

Talevi, A. (2011). *Virtual Screening: An Emergent , Key Methodology for Drug Development in an Emergent Continent-A Bridge Towards Patentability* (pp. 229–245). In *Advanced Methods and Applications in Chemoinformatics: Research*

Progress and New Applications.
<https://www.semanticscholar.org/paper/Virtual-Screening-%3A-An-Emergent-%2C-Key-Methodology-Talevi/b1f67cc3b4b41c0ecdbdfce684a7e7aea31da40e>

Tang, S., Chen, R., Lin, M., Lin, Q., Zhu, Y., Ding, J., Hu, H., Ling, M., & Wu, J. (2022). Accelerating AutoDock Vina with GPUs. *Molecules*, 27(9), 3041. <https://doi.org/10.3390/molecules27093041>

The AlphaFold Team & Google Colab. (2022). *ColabFold*. <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=G4yBrceuFbf3>

Toropova, A. P., & Toropov, A. A. (2017). CORAL: Binary classifications (active/inactive) for drug-induced liver injury. *Toxicology Letters*, 268, 51–57. <https://doi.org/10.1016/j.toxlet.2017.01.011>

Trevethan, R. (2017). Sensitivity, Specificity, and Predictive Values: Foundations, Plausibilities, and Pitfalls in Research and Practice. *Frontiers in Public Health*, 5, 307. <https://doi.org/10.3389/fpubh.2017.00307>

Triballeau, N., Acher, F., Brabet, I., Pin, J.-P., & Bertrand, H.-O. (2005). Virtual Screening Workflow Development Guided by the “Receiver Operating Characteristic” Curve Approach. Application to High-Throughput Docking on Metabotropic Glutamate Receptor Subtype 4. *Journal of Medicinal Chemistry*, 48(7), 2534–2547. <https://doi.org/10.1021/jm049092j>

Tropsha, A. (2010). Best Practices for QSAR Model Development, Validation, and Exploitation. *Molecular Informatics*, 29(6–7), 476–488. <https://doi.org/10.1002/minf.201000061>

Tropsha, A., Gramatica, P., & Gombar, V. (2003). The Importance of Being Earnest: Validation is the Absolute Essential for Successful Application and Interpretation of QSPR Models. *QSAR & Combinatorial Science*, 22(1), 69–77. <https://doi.org/10.1002/qsar.200390007>

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. <https://doi.org/10.1002/jcc.21334>

Truchon, J.-F., & Bayly, C. I. (2007). Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem. *Journal of Chemical Information and Modeling*, 47(2), 488–508. <https://doi.org/10.1021/ci600426e>

Ul-Haq, Z., & D. Madura, J. (Eds.). (2015). How to Judge Predictive Quality of Classification and Regression Based QSAR Models? In *Frontiers in Computational*

Chemistry (pp. 71–120). BENTHAM SCIENCE PUBLISHERS.
<https://doi.org/10.2174/9781608059782115020005>

Valli, M., Dos Santos, R. N., Figueira, L. D., Nakajima, C. H., Castro-Gamboa, I., Andricopulo, A. D., & Bolzani, V. S. (2013). Development of a Natural Products Database from the Biodiversity of Brazil. *Journal of Natural Products*, 76(3), 439–444. <https://doi.org/10.1021/np3006875>

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A. E., & Berendsen, H. J. C. (2005). GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry*, 26(16), 1701–1718. <https://doi.org/10.1002/jcc.20291>

Veerasingam, R., Rajak, H., Jain, A., Sivadasan, S., Varghese, C. P., & Agrawal, R. K. (2011). Validation of QSAR models-strategies and importance. *Int. J. Drug Des. Discov*, 3, 511-519.

Volkamer, A., Griewel, A., Grombacher, T., & Rarey, M. (2010). Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. *Journal of Chemical Information and Modeling*, 50(11), 2041–2052. <https://doi.org/10.1021/ci100241y>

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., & Rarey, M. (2012). Combining Global and Local Measures for Structure-Based Druggability Predictions. *Journal of Chemical Information and Modeling*, 52(2), 360–372. <https://doi.org/10.1021/ci200454v>

Vyškovský, R., Schwarz, D., Janoušová, E., & Kašpárek, T. (2016). *Random Subspace Ensemble Artificial Neural Networks for First-episode Schizophrenia Classification*. 317–321. <https://doi.org/10.15439/2016F333>

Wang, C., & Zhang, Y. (2017). Improving scoring-docking-screening powers of protein-ligand scoring functions using random forest. *Journal of Computational Chemistry*, 38(3), 169–177. <https://doi.org/10.1002/jcc.24667>

Ward, J. H. (1963). Hierarchical Grouping to Optimize an Objective Function. *Journal of the American Statistical Association*, 58(301), 236–244. <https://doi.org/10.1080/01621459.1963.10500845>

Ward, S. E., & Davis, A. M. (Eds.). (2023). *The handbook of medicinal chemistry: Principles and practice* (2nd edition). Royal Society of Chemistry.

Waskom, M. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>

Weininger, D. (1988). SMILES, a chemical language and information system. 1. Introduction to methodology and encoding rules. *Journal of Chemical Information and Computer Sciences*, 28(1), 31–36. <https://doi.org/10.1021/ci00057a005>

Willett, P. (2006). Similarity-based virtual screening using 2D fingerprints. *Drug Discovery Today*, 11(23–24), 1046–1053. <https://doi.org/10.1016/j.drudis.2006.10.005>

Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., Assempour, N., Iynkkaran, I., Liu, Y., Maciejewski, A., Gale, N., Wilson, A., Chin, L., Cummings, R., Le, D., ... Wilson, M. (2018). DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>

WormBase ParaSite. (2023). *Taenia solium*. Retrieved September 13, 2023, from https://parasite.wormbase.org/Taenia_solium_prjna170813/Info/Index/

Yang, Y., Chen, H., Nilsson, I., Muresan, S., & Engkvist, O. (2010). Investigation of the Relationship between Topology and Selectivity for Druglike Molecules. *Journal of Medicinal Chemistry*, 53(21), 7709–7714. <https://doi.org/10.1021/jm1008456>

Zhang, L., Li, D., Xu, R., Zheng, S., He, H., Wan, J., & Feng, Q. (2014). Structural and Functional Analyses of a Sterol Carrier Protein in *Spodoptera litura*. *PLoS ONE*, 9(1), e81542. <https://doi.org/10.1371/journal.pone.0081542>

Zhang, Q., & Muegge, I. (2006). Scaffold Hopping through Virtual Screening Using 2D and 3D Similarity Descriptors: Ranking, Voting, and Consensus Scoring. *Journal of Medicinal Chemistry*, 49(5), 1536–1548. <https://doi.org/10.1021/jm050468i>

Zimmerman, A. W., & Veerkamp, J. H. (2002). New insights into the structure and function of fatty acid-binding proteins. *Cellular and Molecular Life Sciences (CMLS)*, 59(7), 1096–1116. <https://doi.org/10.1007/s00018-002-8490-y>

CAPÍTULO 4: RESULTADOS

4.1 COMPILADO DEL CONJUNTO DE DATOS

Tomando como referencia el conjunto de datos publicado por Bégamo et al. (2020), se amplió la búsqueda bibliográfica obteniendo así un total de 21 trabajos científicos que presentan datos sobre unión de distintos ligandos a **AFABP** (Bégamo et al., 2020; Xu et al., 2012; Okada et al., 2012; Hertzell et al., 2009; Floresta et al., 2019; Yulong et al., 2019; Cai et al., 2017; Floresta et al., 2018; Kühne et al., 2018).

Con esta información, se compiló un conjunto de datos con un total de 288 estructuras químicas, de las cuales 187 se clasificaron como compuestos **ACTIVOS** y 101 como compuestos **INACTIVOS** en términos de su unión a **AFABPs**. Cabe recordar que las moléculas fueron clasificadas en **ACTIVOS** o **INACTIVOS** en función de sus valores de **IC₅₀** ó **Ki** reportadas, donde valores menores a 10 μM fueron considerados **ACTIVOS**, mientras que valores mayores a 20 μM , fueron clasificados como **INACTIVOS**.

La similitud molecular entre todos los compuestos compilados, se visualiza mediante un mapa de calor, utilizando el paquete Seaborn de Python (Waskom, 2021). Para este análisis, se emplearon las huellas digitales moleculares tipo Daylight y el coeficiente de similitud de Tanimoto en su forma binaria, utilizando el paquete RDKit de Python (Rational Discovery LLC, 2006).

El mapa de calor representado en la **Figura 4.1**, muestra los valores del coeficiente de Tanimoto en una escala de color para cada par de compuestos compilados de literatura. Valores más cercanos a 0 (bits azules) indican mayor disimilitud molecular, mientras que regiones más blancas tienden a valores cercanos a 1, indicando mayor similitud estructural entre los compuestos.

El mapa de calor revela que existen pares de compuestos que comparten una alta similitud molecular entre sí, tanto en la comparativa intra como inter categoría. Este último dato (alta similitud entre pares de compuestos de distintas categorías) supone un desafío significativo para los modelos de clasificación, ya que implica que la distinción entre compuestos con capacidad de unión y aquellos sin capacidad de unión puede ser compleja basándose únicamente en sus estructuras moleculares. A pesar de ello, hay un predominio claro de bits azules, lo que significa diversidad en estructuras químicas, y a priori, un dominio de aplicación amplio para los modelos a obtener.

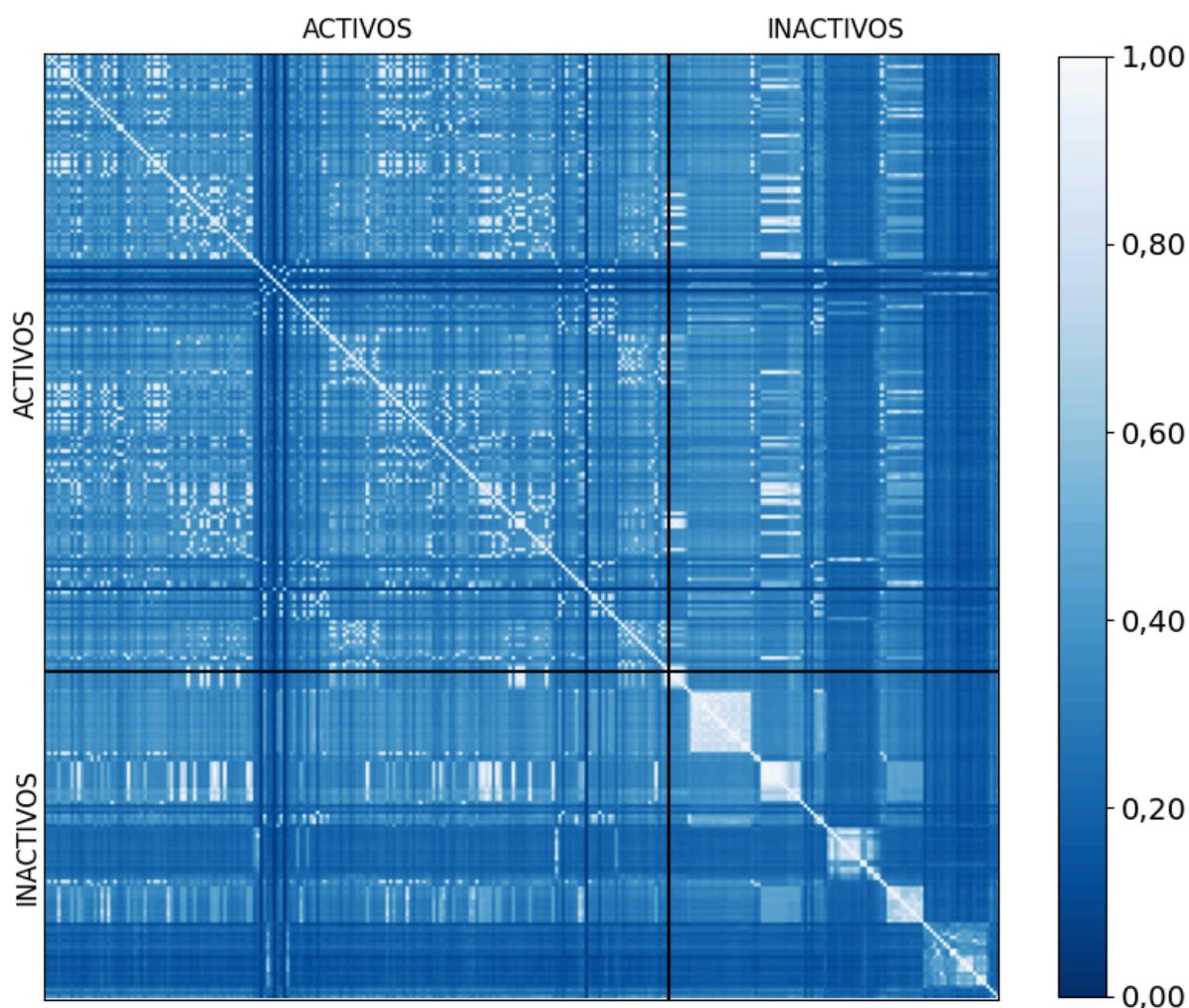


Figura 4.1. Mapa de calor de similitud de los compuestos ACTIVOS e INACTIVOS del conjunto de datos compilado.

4.2 PARTICIÓN DEL CONJUNTO DE DATOS

Como ya se mencionó con anterioridad, el conjunto de datos consiste en 187 compuestos ACTIVOS y 101 compuestos INACTIVOS. Con la finalidad de contar con el conjunto de entrenamiento balanceado con la mayor cantidad de compuestos posible en cada categoría, se optó por utilizar todos los compuestos INACTIVOS dentro de dicho conjunto.

En cuanto a los 187 compuestos ACTIVOS, se aplicó el algoritmo de *clustering* no jerárquico **iRaPCA** (Prada Gori et al., 2022), lo que con una única ronda de *clustering* resultó en cuatro grupos, conteniendo 69, 81, 25 y 12 compuestos, con un valor de **SIL** de 0,8912. Debido a que el conjunto de entrenamiento debe estar balanceado para evitar sesgos hacia una categoría en particular, se seleccionaron compuestos de cada grupo de manera proporcional a fin de llegar a los 101 compuestos ACTIVOS. Los 86 compuestos ACTIVOS remanentes se utilizaron para

la generación de *decoys*, obteniendo así dos quimiotecas simuladas (**LUDe** y **DUDE-Z**), que hicieron las veces de conjuntos de prueba para ejercicios de **CV** retrospectivo.

En la **Tabla 4.2** se presenta la composición final de los conjuntos que se utilizaron para generar y validar los modelos. Además, en el **Material Suplementario 4 (Tabla S4a)**, se proporcionan las estructuras químicas en notación **SMILES** de los compuestos **ACTIVOS** e **INACTIVOS** del conjunto de datos compilado.

Tabla 4.2. Composición de los conjuntos de entrenamiento y de prueba.

	CONJUNTO DE ENTRENAMIENTO	CONJUNTO DE PRUEBA	
		LUDe	DUDE-Z
ACTIVOS	101	86	86
INACTIVOS	101	-	-
DECOYS	-	7613	4965
TOTAL	202	7699	5051

Como puede apreciarse en la **Tabla 4.2**, el porcentaje de activos en la quimioteca **LUDe** es de aproximadamente el 1,1%, mientras que en **DUDE-Z** es de 1,7%. La finalidad de estas quimiotecas simuladas, como ya se mencionó en el capítulo anterior, es la de evaluar la capacidad predictiva de los modelos generados en una quimioteca con una baja proporción de compuestos activos, situación que se asemeja a la de un **CV**. Vale recordar que cuando la proporción de compuestos activos en las quimiotecas simuladas es menor a 0,05, la desviación estándar de **AUCROC** converge a un valor constante, lo cual proporciona resultados más robustos estadísticamente (Truchon & Bayly, 2007).

4.3 GENERACIÓN Y VALIDACIÓN DE MODELOS

Se generaron 4000 modelos lineales a partir de la misma cantidad de subconjuntos aleatorios, en donde cada conjunto incluyó un máximo de 250 descriptores. Estos modelos clasificatorios buscan detectar compuestos capaces de unirse a las **FABPs** de cestodos.

Los modelos se ordenaron de manera decreciente en función de su valor de **AUCROC** en la quimioteca simulada **LUDe**. Por otro lado, si dos modelos comparten más de un descriptor molecular, se elimina de la lista aquel modelo del par con menor **AUCROC**. De esta manera, se evitará, en la instancia de ensamblado de modelos, la combinación de modelos altamente redundantes.

La **Ecuación 4.1** representa al **Modelo 3963**, el modelo individual obtenido con el mejor desempeño:

$$Y = 0,3832 * nAcid - 0,0399 * ATSC8are - 0,0211 * AATS6i - 0,2626 * AATSC8i + 0,0014 * ATSC3dv + 3,5257 \quad (4.1)$$

donde, *nAcid* representa a la cantidad de grupos ácidos que contiene la molécula, *ATSC8are* se corresponde con el coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por electronegatividad de Allred-Rocow, *AATS6i* es el coeficiente de autocorrelación promediado de Broto-Moreau de orden 6 ponderado por potencial de ionización, *AATSC8i* se refiere al coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 8 ponderado por potencial de ionización y *ATSC3dv* al coeficiente de autocorrelación centrado de Broto-Moreau de orden 3 ponderado por electrones de valencia. Muy brevemente, un descriptor de autocorrelación 2D es aquel que cuantifica cómo una propiedad determinada se distribuye a lo largo de la topología de una estructura molecular (Pharmacognosy group, Uppsala University, 2023). De esta manera, el orden del coeficiente indica a cuántos enlaces de distancia se encuentran los átomos entre los que se está haciendo la comparativa de la propiedad molecular en cuestión. Una explicación más detallada de este tipo de descriptores se puede encontrar en el **Material Suplementario 4**.

Como se puede observar en la **Ecuación 4.1**, valores mayores del descriptor *nAcid* se corresponden con un mayor valor de puntaje para un molécula dada (*Y*). De esta observación puede decirse que, dentro del rango de valores abarcado por el conjunto de entrenamiento, a mayor cantidad de grupos ácidos contenga la molécula, más probabilidades tiene de ser un compuesto clasificado como ACTIVO por el modelo, lo cual parece estar en consonancia con la función postulada para las FABPs como transportadoras de ácidos grasos. Una situación similar ocurre con el descriptor *ATSC3dv*, aunque la interpretación para este descriptor no es directa.

Por otra parte, los tres descriptores restantes incluyen a propiedades como la electronegatividad y al potencial de ionización, donde valores mayores en estas características se relacionan con valores más bajos de *Y*. Si se piensa en moléculas hidrofóbicas, uno puede pensar que la cantidad de átomos electronegativos presentes en ellas es muy pequeña en relación al total de átomos que las componen, con lo cual esta relación tendría sentido. Por otra parte, la presencia de grupos ácidos ionizables permite pensar que los valores del potencial de ionización serían más bien bajos en comparación con moléculas sin grupos ionizables, explicando la relación inversa entre potencial de ionización e *Y*.

Los modelos ordenados y filtrados por descriptores se ensamblaron para evaluar su desempeño en relación al **Modelo 3963**. Para ello, se utilizaron los operadores **PROD**, **MIN**, **PROM**, **RANK** y **VOT** y la métrica **BEDROC** para la evaluación de cada uno de ellos, buscando el ensamble de modelos que maximice la mencionada métrica. La variación del valor de **BEDROC** en función de los modelos ensamblados para cada operador se llevó adelante con un subconjunto de la quimioteca **LUDe** (**LUDe-A**, 43 activos y 3807 señuelos) mientras que el resto de la quimioteca fue utilizada para la evaluación del ensamble seleccionado (**LUDe-B**, 43 activos y 3806 señuelos). Los resultados se muestran en la **Figura 4.3** y **Tabla 4.3a**.

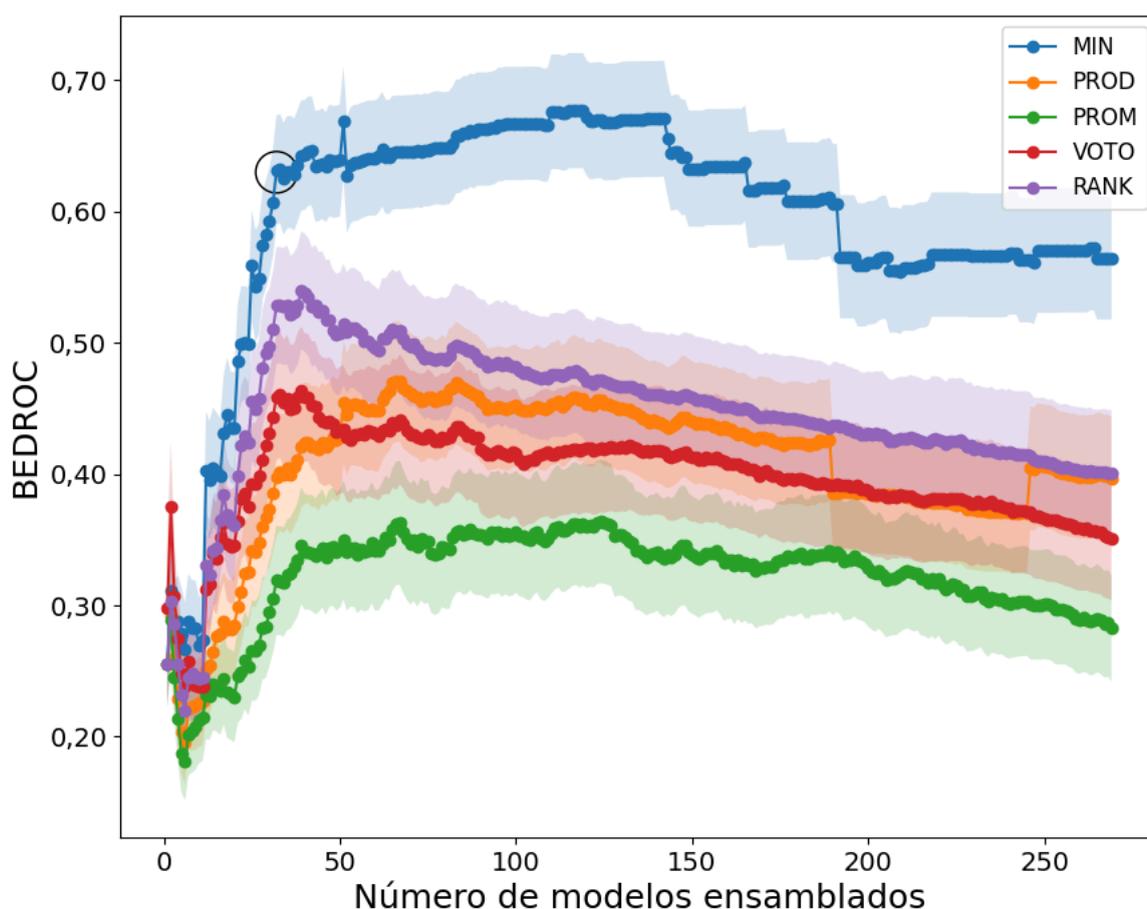


Figura 4.3. Variación de la métrica **BEDROC** ($\alpha=100$) en función del número de modelos ensamblados para cada uno de los operadores ensayados. El espacio sombreado para cada curva se corresponde con el desvío estándar para cada punto, el cual fue obtenido por *bootstrapping*. El ensamble **MIN32** se encuentra indicado por un círculo.

Tabla 4.3a. Valores de **AUCROC** y **BEDROC** para el mejor modelo individual (**Modelo 3963**) y para el ensamble **MIN32**.

MODELO	LUDe-A AUCROC (+/- DE)	LUDe-A BEDROC (+/- DE)	LUDe-B AUCROC (+/- DE)	LUDe-B BEDROC (+/- DE)
3963	0,932 (+/- 0,015)	0,255 (+/- 0,032)	0,954 (+/- 0,005)	0,273 (+/- 0,038)
MIN32	0,945 (+/- 0,015)	0,632 (+/- 0,041)	0,935 (+/- 0,013)	0,603 (+/- 0,034)

Como se puede observar en la **Figura 4.3**, el operador **MIN** presenta el mejor desempeño en términos de la métrica **BEDROC**, y teniendo en cuenta el número de modelos ensamblados, los mejores 32 modelos individuales presentan un valor máximo en términos de la métrica evaluada. Si bien ensambles con mayor cantidad de modelos pueden presentar valores similares, el ensamble **MIN32** presenta un desempeño similar al de otros ensambles con menor cantidad de modelos involucrados, por lo cual, siguiendo el principio de parsimonia, se lo eligió para el **CV** prospectivo. De la misma manera, se puede notar cómo mejora el desempeño en términos de **BEDROC** para el ensamblado de modelos en relación al mejor modelo individual (**Tabla 4.3a**). Debido a que el valor de α utilizado fue de 100, la probabilidad de que un compuesto ACTIVO se encuentre en el 1,6% mejor rankeado de las quimiotecas **LUDe-A** y **LUDe-B** es algo más de 0,6.

Una vez seleccionado el ensamble, se procedió a estimar la capacidad predictiva del mismo mediante las validaciones internas. Como se mencionó en el capítulo 3 de esta tesis doctoral, se llevaron adelante las metodologías de aleatorización de Fisher y **LGO**, para evaluar la robustez de los 32 modelos en el conjunto de entrenamiento. Para la implementación de **LGO** se removieron 20 compuestos del conjunto de entrenamiento por ronda (10 ACTIVOS y 10 INACTIVOS, 1000 rondas), mientras que con los ejemplos restantes se generaron los 32 modelos que conforman el ensamble **MIN32**. Los resultados se muestran en la **Tabla 4.3b**.

Tabla 4.3b. Resultados de la validación interna para los 32 modelos que conforman el ensamble *MIN32*. El porcentaje de buenas clasificaciones (%BC) representa la media para 1000 rondas de aleatorización para cada caso.

NÚMERO DE MODELO	MODELO ORIGINAL	ALEATORIZACIÓN DE FISHER		LEAVE GROUP OUT	
	%BC	%BC	DESVÍO ESTÁNDAR	%BC	DESVÍO ESTÁNDAR
3963	79,70	49,38	11,30	71,87	9,40
1940	80,20	50,00	12,05	73,02	9,24
2836	77,72	50,00	12,50	72,53	9,33
3392	80,20	50,00	13,15	73,88	9,42
3555	81,19	50,00	13,46	73,57	9,78
498	78,71	50,00	12,70	75,04	9,41
3520	80,69	50,00	13,33	74,53	8,92
2525	74,26	50,00	11,74	69,95	9,21
3732	76,73	50,00	11,90	72,30	9,42
3668	76,73	50,00	13,07	75,14	9,08
2808	77,23	50,00	12,45	71,03	9,10
2892	74,75	50,00	12,19	70,44	9,40
527	75,25	50,00	12,06	70,33	9,55
1992	78,22	50,00	13,83	75,40	9,22
2346	72,77	50,00	11,55	68,65	9,74
3292	75,74	49,00	14,39	71,09	9,25
1412	77,72	49,00	11,83	74,05	9,02
3674	74,75	50,00	12,08	70,80	10,20
2689	76,73	50,00	11,59	69,59	9,87
3558	77,23	49,00	13,05	74,11	8,93
3896	74,75	49,00	11,00	66,44	9,57
2933	73,27	49,00	11,73	67,10	9,85
3744	76,73	49,00	11,51	69,85	10,19
3390	73,76	50,00	11,86	69,81	9,56
769	74,26	50,00	11,71	70,00	9,27
2424	74,26	50,00	12,75	69,62	9,72

327	73,76	50,00	11,86	69,81	9,56
2269	76,73	50,00	12,31	70,94	9,64
3976	77,23	50,00	11,86	70,65	9,85
3380	76,24	49,00	11,99	71,47	9,70
1358	76,73	49,00	12,08	72,24	9,36
1178	72,77	50,00	14,06	70,61	9,73

Como se aprecia en la **Tabla 4.3b**, los **%BC** son muy similares a los obtenidos por los modelos originales, con las diferencias en cada caso dentro del desvío estándar de las aleatorizaciones.

En cuanto a la aleatorización de Fisher, se observa que en todos los casos el rendimiento promedio de los modelos generados al azar es significativamente inferior al rendimiento de los modelos originales, lo que sugiere una baja probabilidad de correlaciones espurias. En el **Material Suplementario 4 (Lista S4, Tabla S4b y Ecuaciones S4.1-S4.3)** se muestran las expresiones matemáticas de los 32 modelos del ensamble **MIN32**, junto con una descripción breve de cada uno de los descriptores incluidos en los mismos.

4.4 SELECCIÓN DEL PUNTO DE CORTE PARA EL CRIBADO

La utilización del operador **MIN** para combinar modelos resulta en los logros más sobresalientes en términos de la **Sp**. Este aspecto es de gran importancia, dado que reduce significativamente el número de falsos positivos, es decir, compuestos seleccionados como activos por el modelo pero que no presentan actividad *in vitro* contra nuestra diana farmacológica. Optimizar la **Sp** juega un papel crucial en el enfoque de esta tesis, la cual se desarrolló en laboratorios públicos con recursos limitados para la adquisición y evaluación de candidatos. Por otra parte, se espera que la cantidad de compuestos inactivos en las quimiotecas/bases de datos tamizadas supere ampliamente a los compuestos activos, razón por la cual modelos aún con muy buen desempeño clasificatorio pueden seleccionar una cantidad relativamente alta de falsos positivos. Por lo tanto, al priorizar la **Sp**, se minimiza la probabilidad de seleccionar falsos positivos para la etapa experimental de evaluación, a expensas de perder **Se** en el tamizado, es decir perder potenciales compuestos **ACTIVOS** para ensayar porque el ensamble los categoriza como **INACTIVOS**.

A fin de optimizar la **Sp** por sobre la **Se**, en función de la discusión previa, se analizaron las superficies de **PPV**, para las quimiotecas simuladas **LUDe-B** y **DUDE-Z**. Cabe recordar que el valor numérico de **PPV** determina la capacidad para predecir correctamente las instancias positivas, es decir, cuántos de esos

compuestos seleccionados como **ACTIVOS** por el modelo confirmarán tal predicción en la instancia de determinación experimental.

El valor de corte seleccionado para la campaña de **CV** fue de 0,400. Para dicho valor de corte, tenemos un valor de 0,994 para **Sp**, de 0,455 para **Se** y de 0,455 para **PPV** ($Y_a=0,01$). El valor de **PPV** indica la probabilidad que tiene el compuesto seleccionado por el ensamble de presentar actividad contra la diana farmacológica, si el porcentaje de compuestos activos es del 1% en la base de datos cribada.

Como validación adicional, se analizó el valor de **PPV** para **DUDE-Z** utilizando el valor de corte recién mencionado. Los valores de **Sp**, **Se** y de **PPV** ($Y_a=0,01$) fueron 0,995, 0,455, y 0,496, respectivamente. Para esta quimioteca simulada, la probabilidad de encontrar un compuesto activo experimental seleccionado por el ensamble es de casi 5 cada 10, considerando un 1% de activos. Este resultado es similar al obtenido para **LUDe-B**. Las **Figuras 4.4a,b** y la **Tabla 4.4** muestran los resultados.

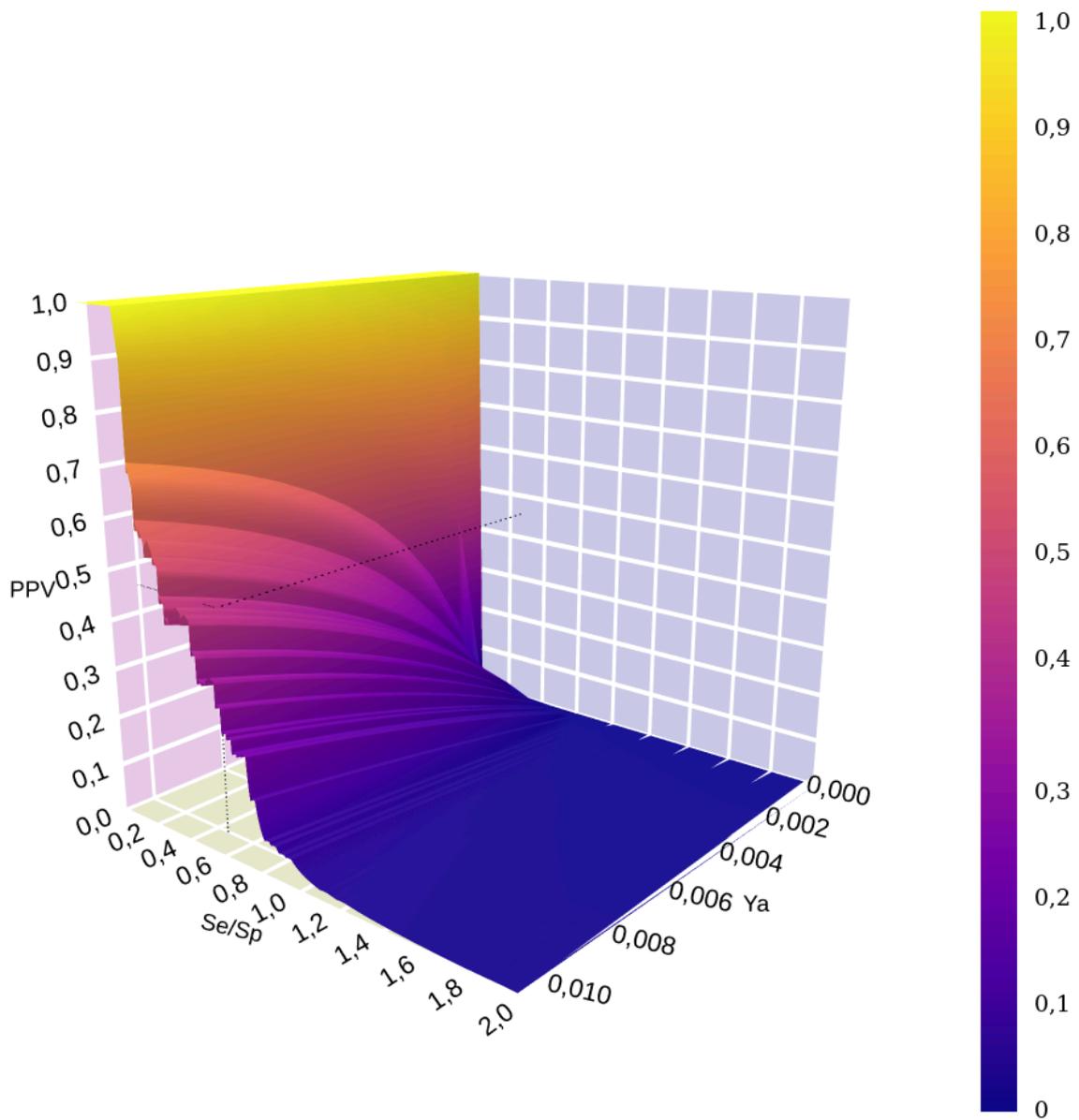


Figura 4.4a. Superficie de **PPV** para el ensamble **MIN32** con la quimioteca **LUDe-B**. Las líneas punteadas señalan las coordenadas del valor de corte seleccionado en este trabajo de tesis doctoral.

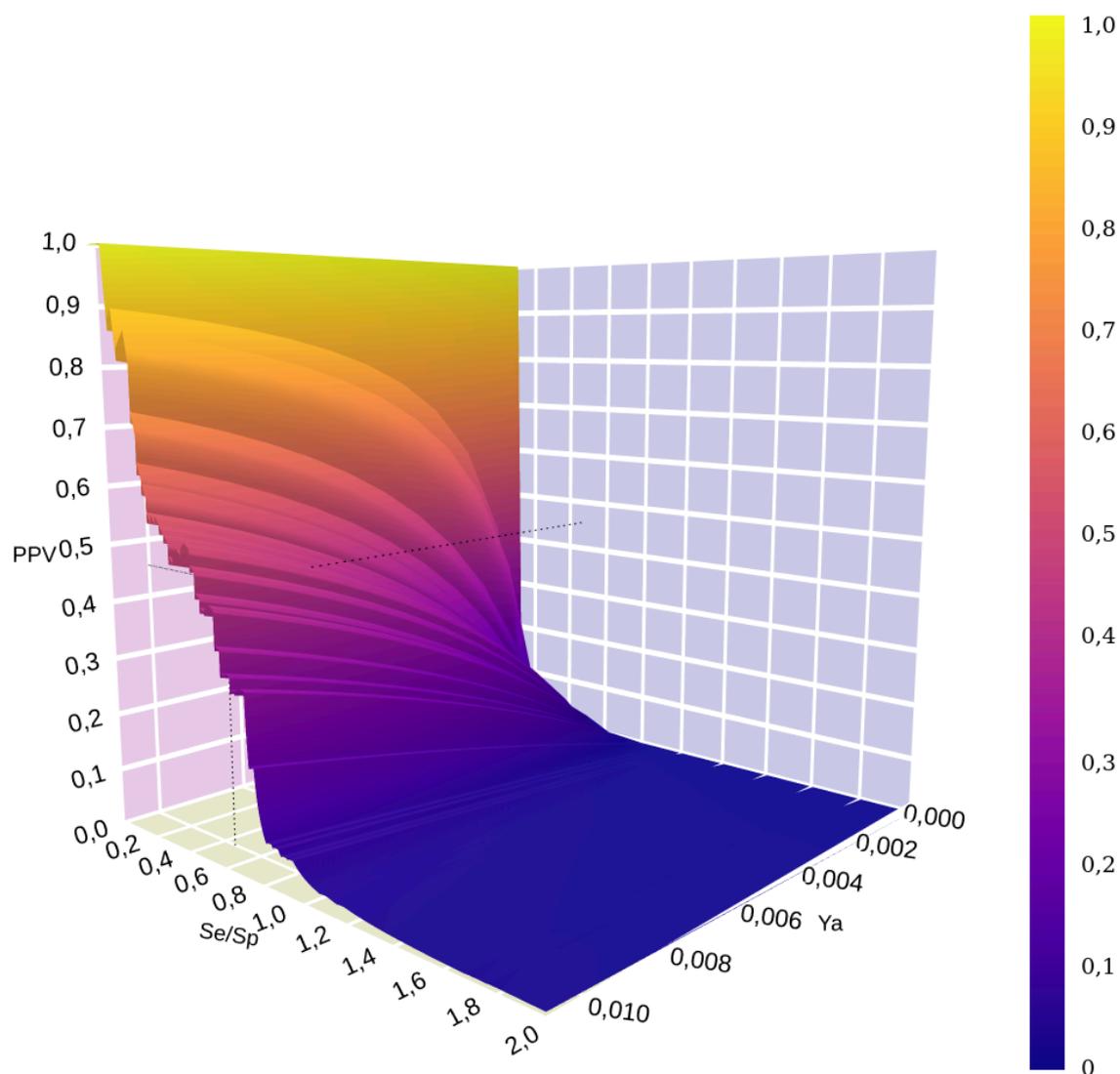


Figura 4.4b. Superficie de **PPV** para el ensamble **MIN32** con la quimioteca **DUDE-Z**. Las líneas punteadas señalan las coordenadas del valor de corte seleccionado en este trabajo de tesis doctoral.

Tabla 4.4. Valores de **Se**, **Sp** y **PPV** para el valor de corte establecido para el **CV** prospectivo. Los valores entre paréntesis indican la proporción de compuestos activos considerada para el cálculo de **PPV**.

QUIMIOTECA	VALOR DE CORTE	Se	Sp	PPV (0,001)	PPV (0,005)	PPV (0,01)
LUDe-B	0,400	0,455	0,994	0,076	0,294	0,455
DUDE-Z	0,400	0,455	0,995	0,089	0,328	0,496

4.5 CRIBADO VIRTUAL

Se aplicó la combinación de modelos **MIN32** para el **CV** prospectivo de las bases de datos DrugBank, DRH, Sweetlead, HybridMolDB, NuBBEDB, NaturAr, COCONUT y LIDeB, contabilizando un total de aproximadamente 435.000 compuestos. Como resultado, se seleccionaron 92 compuestos que fueron identificados como **ACTIVOS**.

Dentro de esos compuestos activos, 15 de ellos se encuentran en algún estadio de ensayos clínicos para su aplicación al tratamiento de otras patologías. En función de su accesibilidad en términos de costo y disponibilidad para su adquisición, se seleccionaron 3 de los 15 compuestos que habían sido predichos como **ACTIVOS** y que pertenecían al dominio de aplicación de los modelos. Estos fármacos fueron adquiridos para su posterior evaluación experimental: Naratriptán HCl (*Cayman Chemical*), Nitrate de Fenticonazol (*Cayman Chemical*), e Hidroclorotiazida (gentilmente donado por Laboratorios Bagó). El Naratriptán es un compuesto que se utiliza para tratar la migraña, mientras que el Fenticonazol es un agente antifúngico de aplicación tópica y la Hidroclorotiazida es un fármaco que actúa como diurético en el tratamiento de la hipertensión arterial e insuficiencia cardíaca. En la **Figura 4.5** se pueden apreciar las estructuras químicas de los tres compuestos seleccionados.

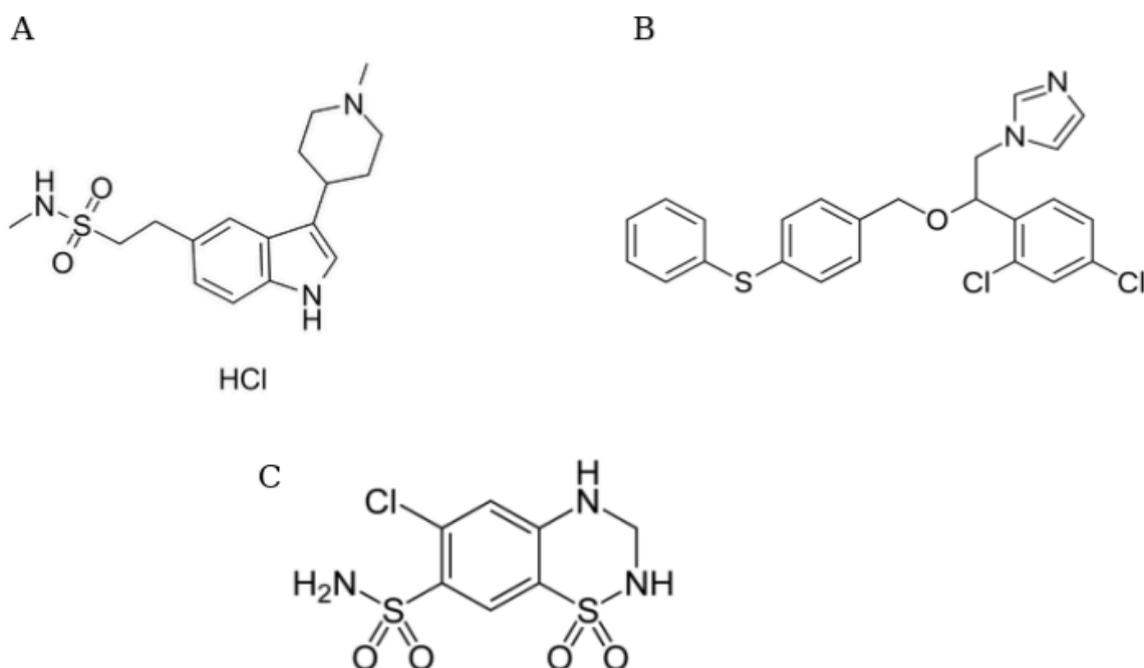


Figura 4.5. Estructuras químicas del Naratriptán HCL (A), Fenticonazol (B) e Hidroclorotiazida (C).

4.6 MODELADO ESTRUCTURAL DE LAS FABPs

Como se mencionó en el capítulo anterior, el modelado estructural de las **FABPs** se llevó adelante utilizando ColabFold (The AlphaFold Team & Google Colab, 2022). Para cada una de las isoformas, el método predijo 5 estructuras posibles, las cuales fueron rankeadas de acuerdo al **score LDDT** (Mariani et al., 2013). Para los fines de este trabajo, solo la estructura mejor rankeada para cada isoforma fue considerada. En la **Figura 4.6** se muestra el **LDDT score** para cada residuo (**pLDDT**) y para cada una de las proteínas modeladas.

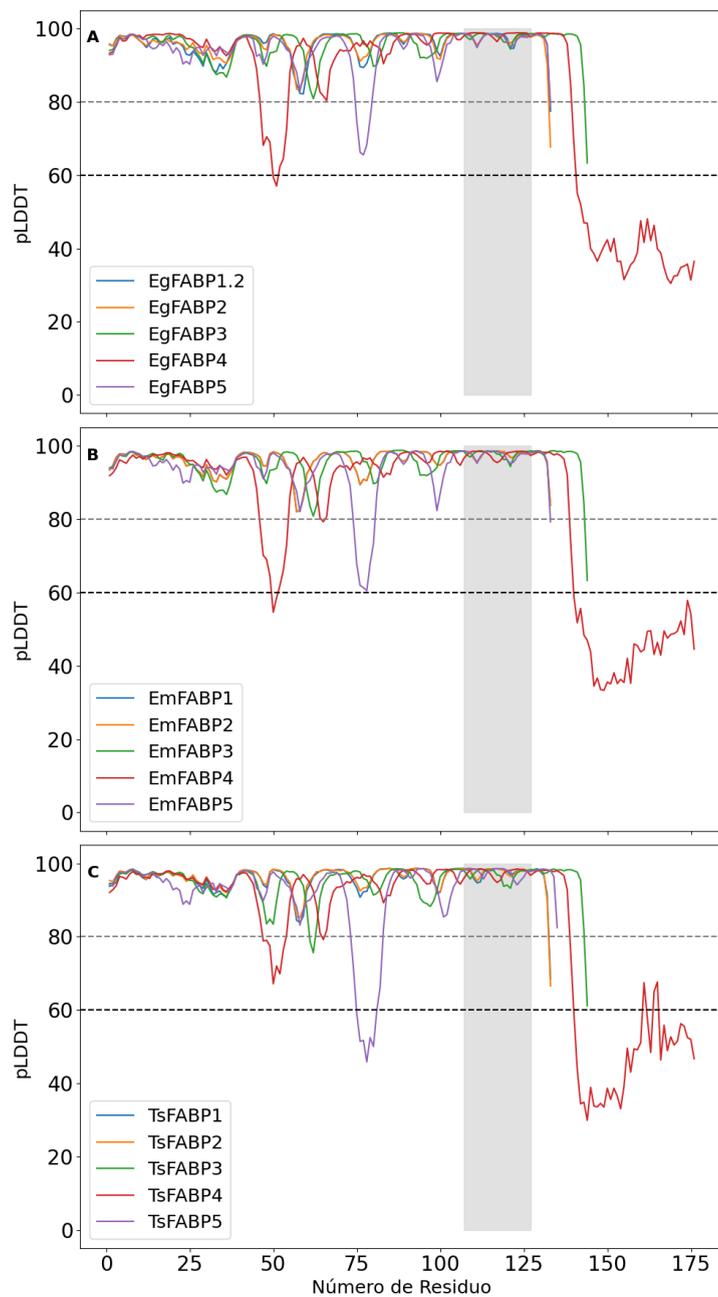


Figura 4.6. pLDDT score por residuo para las isoformas de (A) *E. granulosus* (B) *E. multilocularis* y (C) *T. solium*. Puntajes de pLDDT mayores a 80 indican que el

modelo es confiable en la predicción (línea gris punteada), mientras que valores inferiores a 60 indican que la confiabilidad de la predicción es baja (línea negra punteada). El rectángulo gris indica el sitio de unión reportado para las **FABPs**.

Según se observa en la **Figura 4.6**, casi todos los residuos en todas las isoformas cuentan con un **pLDDT score** superior a 80, lo cual demuestra que la predicción es confiable. Sin embargo, se puede notar que las isoformas 4 en las tres especies de parásitos (**EgFABP4**, **EmFABP4** y **TsFABP4**) presentan una región de baja confianza en el modelado, entre los residuos 130 al 176 aproximadamente, siendo modelada como un elemento sin estructura secundaria asociada (en inglés, *random-coil*). Este hecho, sumado a la observación descrita en el trabajo de Bélgameo et al. (2020), en donde se describe a **EmFABP4** como un dímero en solución, hace difícil un modelado confiable para posterior uso en *docking* molecular, con lo que se decidió abandonar la modelización de estas tres isoformas.

Por otro lado, las isoformas 5 presentan bajos **pLDDT scores** alrededor del aminoácido 75, aunque esto último no representa un mayor problema, debido a que dicha región no se encuentra comprometida en el sitio de unión postulado para las **FABPs** (rectángulo gris en la **Figura 4.6**).

De esta manera, las restantes isoformas modeladas fueron sometidas a **DM** y la posterior generación de ensamblajes conformacionales para su posterior utilización en la campaña de cribado por métodos directos.

4.7 GENERACIÓN Y SELECCIÓN DE CONFORMACIONES POR DINÁMICA MOLECULAR, PCA Y K-MEANS

Los modelos estructurales de las distintas isoformas de las **FABPs** de parásitos fueron simuladas mediante **DM**, como se mencionó en el apartado anterior y en el capítulo 3. Estas simulaciones permiten enfrentar a los modelos obtenidos a un entorno más similar al real, es decir, con moléculas de agua e iones explícitos que influyen en la estructura de las proteínas. De esta manera, se busca la exploración del espacio conformacional de cada isoforma, con el fin de encontrar conformaciones alternativas a la propuesta por el modelado inicial, las cuales serán de utilidad para el *ensemble docking*, estrategia de **CV** utilizada en este trabajo.

Una vez transcurridos los 100 ns de simulación para cada isoforma, se realizó el análisis de **PCA** seguido del *clustering* por **k-means**, obteniendo así un número óptimo de grupos o *clusters* por isoforma, esto último de determinado por la métrica **SIL**. Finalmente, para cada isoforma, se extrajeron el número correspondiente de conformaciones correspondientes al valor máximo de **k**. Las conformaciones seleccionadas fueron aquellas que se encontraban más cercanas a cada centroide del grupo correspondiente.

Los resultados se muestran en la **Figura 4.7** y la **Tabla 4.7**.

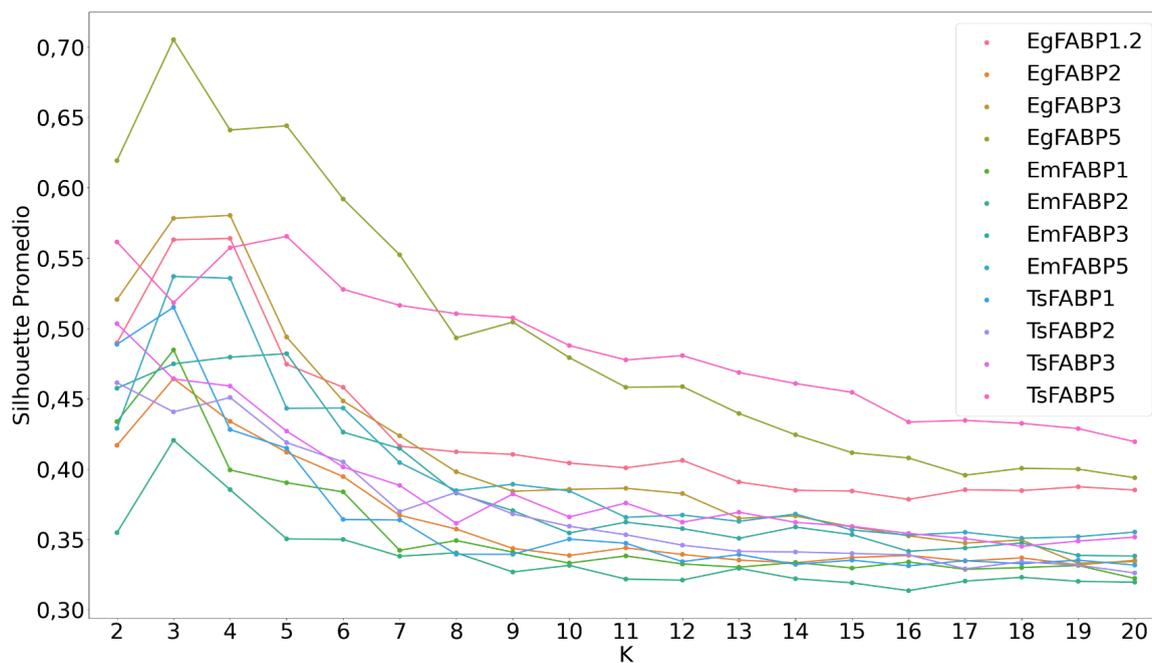


Figura 4.7. Valores de **SIL** promedio para valores de **k** entre 2 a 20 para cada una de las isoformas simuladas por **DM**.

Tabla 4.7. Valores de **SIL** promedio máximo para cada una de las isoformas, junto con su valor de **k**-asociado.

ISOFORMA	SIL PROMEDIO MÁXIMO	k
EgFABP1.2	0,56	4
EgFABP2	0,46	3
EgFABP3	0,58	4
EgFABP5	0,71	3
EmFABP1	0,48	3
EmFABP2	0,42	3
EmFABP3	0,48	5
EmFABP5	0,54	3
TsFABP1	0,52	3
TsFABP2	0,46	2

TsFABP3	0,50	2
TsFABP5	0,57	5

4.8 OBTENCIÓN DE SITIOS DE HIDRATACIÓN POR DINÁMICA MOLECULAR

De la misma manera que se procedió con los modelos, se sometió a **DM** la isoforma **EgFABP1** (PDB:1O8V), en ausencia de su ligando, en pos de encontrar potenciales sitios de hidratación (**WS**) con el algoritmo de WATCLUST (López et al., 2015).

El análisis de WATCLUST sobre la trayectoria de **EgFABP1** evidenció la existencia de tres *clusters* de agua cerca de la tríada de unión de las **FABPs** (Jakobsson et al., 2003) (Arg107, Arg127 y Tyr129), la cual se encuentra próxima a moléculas de agua presentes en el cristal. Los parámetros **WFP** (del inglés *Water Finding Probability*, indicativo de la probabilidad de encontrar una molécula de agua en el *cluster* *i* con respecto al agua del seno de la solución (*bulk*)) y **R90** (distancia en la cual se encuentran el 90% de las moléculas de agua del *cluster*, en Å) muestran que la probabilidad de encontrar una molécula de agua en cada uno de dichos *clusters* es aproximadamente cinco veces mayor que en el *bulk*, y por otra parte, que estos *clusters* son densos. De esta manera, se utilizaron las coordenadas del centro de los *clusters* obtenidos para la utilización del algoritmo de *docking* AutoDock Bias (Arcon et al., 2019), descrito en el capítulo 3.

Los resultados se ilustran en la **Figura 4.8** y la **Tabla 4.8**.

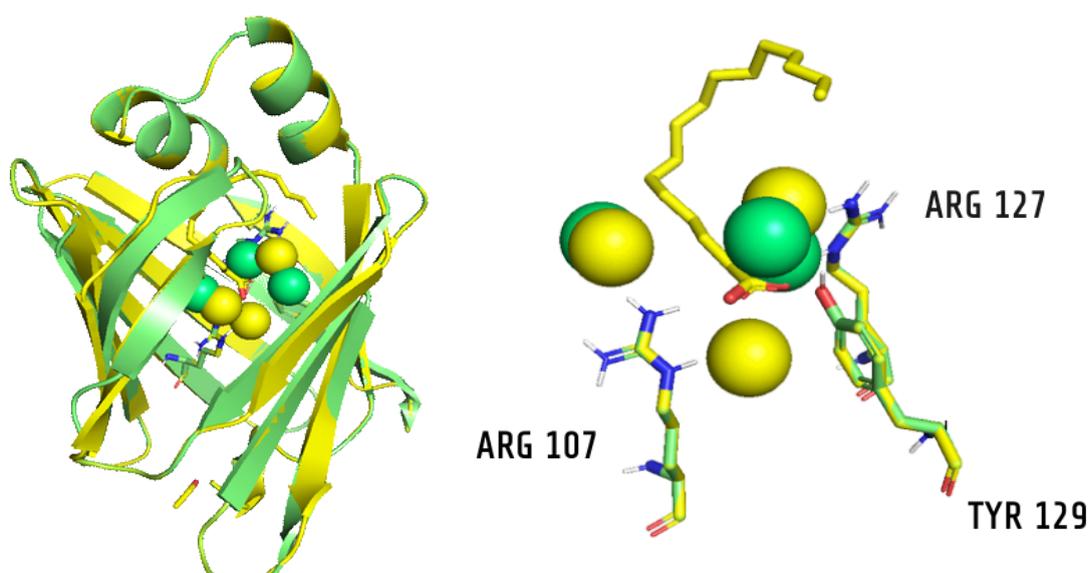


Figura 4.8. Estructura del cristal de **EgFABP1** (amarillo) en superposición con el resultado de la simulación por **DM** (verde). Se pueden ver a los *clusters* de agua

predichos por WATCLUST en comparación con las moléculas presentes en el cristal, junto con la tríada de las **FABPs** (lado derecho de la figura).

Tabla 4.8. Parámetros para los tres clusters de agua predichos para **EgFABP1**.

CLUSTER	WFP	R90 (Å)	NÚMERO DE MOLÉCULAS
1	5,62	0,80	1856
2	4,83	1,00	1824
3	4,95	1,20	1584

4.9 VALIDACIÓN DE LA FUNCIÓN DE PUNTUACIÓN

Se llevó adelante la validación de la función de puntuación (*scoring power*) de los diferentes algoritmos de *docking* utilizados en esta tesis doctoral: **Vina** (Trott & Olson, 2010), **Vina + Vinardo** (Quiroga & Villarreal, 2016), **AD GPU** (Santos-Martins et al., 2021), **AD Hyd** (Forli & Olson, 2012) y **AD Bias** (Arcon et al., 2019). Dado que se observan tres **WS**, y que las moléculas de agua pueden actuar como dadoras y/o aceptoras de puentes de hidrógeno, para el caso de **AD Bias** se llevaron adelante todas las combinaciones posibles de donador/aceptor entre estos tres sitios, denotados por una letra D o A, respectivamente. En esta validación, se utilizó la estructura cristalina de **EgFABP1**, mencionada anteriormente. Para el cálculo de **AUCROC** se consideró el puntaje de la mejor de las poses predichas por cada programa.

El algoritmo que consiguió el mejor desempeño en la validación del *scoring power* fue **AD Bias**, estando el valor de **AUCROC** para cualquiera de sus variantes por encima de los algoritmos restantes, en ambas quimiotecas simuladas. A su vez, cuando los tres **WS** actúan como donores de puentes de hidrógeno (**AD Bias DDD**), se observa el máximo desempeño en términos de **AUCROC**, en comparación con las demás variantes propuestas para **AD Bias**. Los resultados se muestran en la **Tabla 4.9a** y **Figura 4.9a**.

Tabla 4.9a. Valores de **AUCROC** para las quimiotecas **LUDe** y **DUDE-Z** para cada uno de los programas utilizados. *Sin diferencias significativas en Test de Tukey ($\alpha=0,05$). #Diferencias significativas en Test de Tukey ($\alpha=0,05$).

ALGORITMO	AUCROC LUDe (+/- DE)	AUCROC DUDE-Z (+/- DE)
Vina	0,6624 +/- 0,0095	0,8016 +/- 0,0070
Vina + Vinardo	0,6923 +/- 0,0074	0,7736 +/- 0,0061
AD GPU	0,6759 +/- 0,0087	0,8265 +/- 0,0068
AD Hyd	0,6589 +/- 0,0096	0,8361 +/- 0,0066
AD Bias (AAA)	0,7054 +/- 0,0096	0,8402 +/- 0,0068
AD Bias (DDD)	0,7670* +/- 0,0080	0,9030# +/- 0,0059
AD Bias (AAD)	0,7218 +/- 0,0085	0,8547 +/- 0,0063
AD Bias (ADA)	0,7497 +/- 0,0084	0,8804 +/- 0,0065
AD Bias (DAA)	0,7368 +/- 0,0093	0,8592 +/- 0,0066
AD Bias (ADD)	0,7513 +/- 0,0084	0,8838 +/- 0,0065
AD Bias (DAD)	0,7334 +/- 0,0089	0,8632 +/- 0,0064
AD Bias (DDA)	0,7676* +/- 0,0083	0,8958# +/- 0,0062

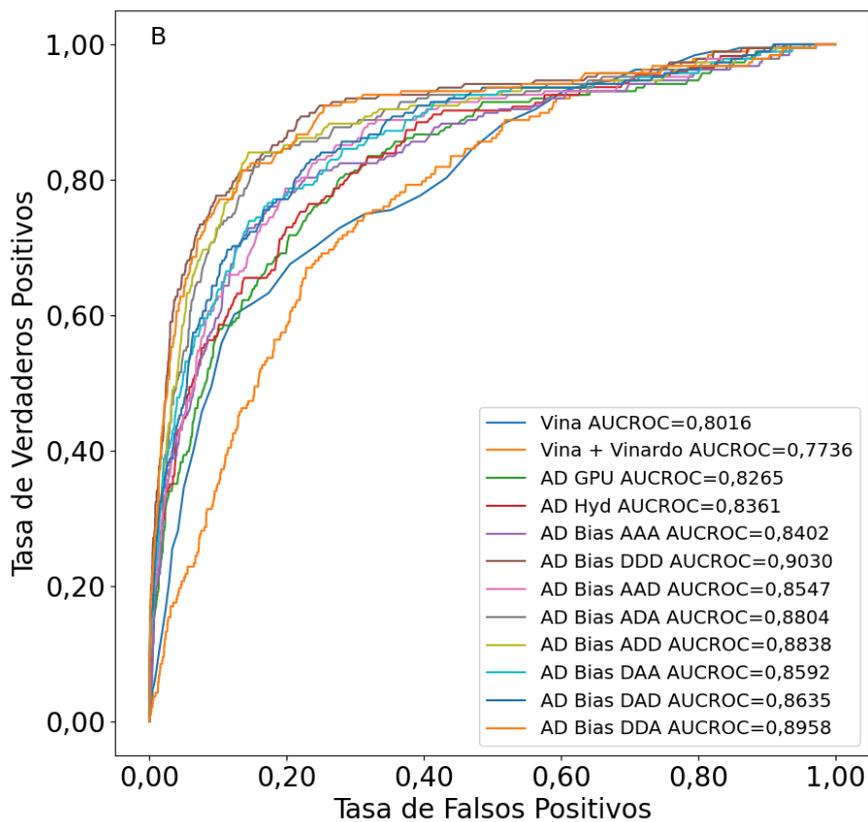
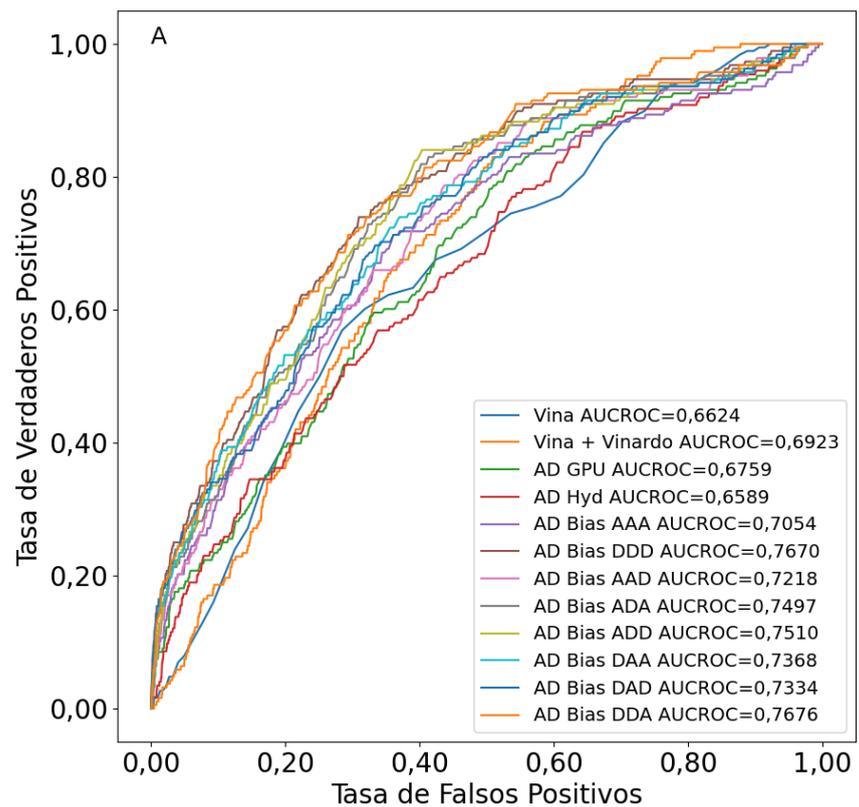


Figura 4.9a. Curvas ROC para cada uno de los algoritmos ensayados en este trabajo, para las quimiotecas **LUDe** (A) y **DUDE-Z** (B).

Una vez seleccionado el algoritmo **AD Bias DDD**, se decidió investigar si los tres **WS** tienen el mismo impacto en el desempeño del *scoring power*. Para ello, se analizaron los **AUCROC** presentados en la **Tabla 4.9a** y cómo éstos varían en función de la posición del sitio donor/aceptor.

Cuando el sitio aceptor (A) es cambiado por un sitio donor (D) para el *cluster 2* (**Tabla 4.8**), se observa el mayor aumento en la **AUCROC** en ambas quimiotecas simuladas. Esta tendencia se observa en menor medida para el sitio del *cluster 1* y finalmente, el cambio es menos importante para el *cluster 3*. Al llevar adelante el análisis inverso, es decir reemplazar un sitio D por uno A a la vez se puede observar la misma tendencia, cuando la sustitución es en el *cluster 2*, la caída en **AUCROC** es mayor, seguida por el reemplazo en el *cluster 1* y finalmente en el 3. Este análisis se muestra en la **Tabla 4.9b**.

Tabla 4.9b. Impacto del sesgo aceptor/donor en función de su posición.

A->D			D->A		
SITIOS	AUCROC LUDe	AUCROC DUDE-Z	SITIOS	AUCROC LUDe	AUCROC DUDE-Z
AAA	0,7054	0,8402	DDD	0,7670	0,9030
AAD	0,7218	0,8547	DDA	0,7676	0,8958
AA <u>A</u>	0,7054	0,8402	D <u>DD</u>	0,7670	0,9030
A <u>DA</u>	0,7497	0,8804	D <u>AD</u>	0,7334	0,8632
<u>AAA</u>	0,7054	0,8402	<u>DDD</u>	0,7670	0,9030
<u>DAA</u>	0,7368	0,8592	<u>ADD</u>	0,7513	0,8838

Debido al impacto del sesgo en cada uno de los sitios en el desempeño del *scoring power*, se decidió ensayar removiendo dichos sesgos donores de manera secuencial, es decir analizar el **AUCROC** quitando el sesgo donor en el *cluster 3* (situación **DD-**) y en el *cluster 1* (**-D-**).

Como se puede observar en la **Tabla 4.9c**, el desempeño del modelo de *docking* mejora en ambas bases de datos cuando se tiene un único sesgo en el *cluster 2*. Por esta situación, sumada al hecho de que el modelo **-D-** es el más sencillo en términos de parámetros, se optó por utilizarlo para la campaña de **CV** prospectivo. La ubicación del sesgo donor en el sitio de unión se visualiza en la **Figura 4.9b**.

Tabla 4.9c. AUCROC para el *scoring power* utilizando 1, 2 o 3 sesgos donores, en las quimiotecas simuladas **LUDe** y **DUDE-Z**.

AUCROC	LUDe (+/- DE)	DUDE-Z (+/- DE)
-D-	0,7794 +/- 0,0077	0,9052 +/- 0,0056
DD-	0,7696 +/- 0,0079	0,9011 +/- 0,0058
DDD	0,7670 +/- 0,0080	0,9030 +/- 0,0059

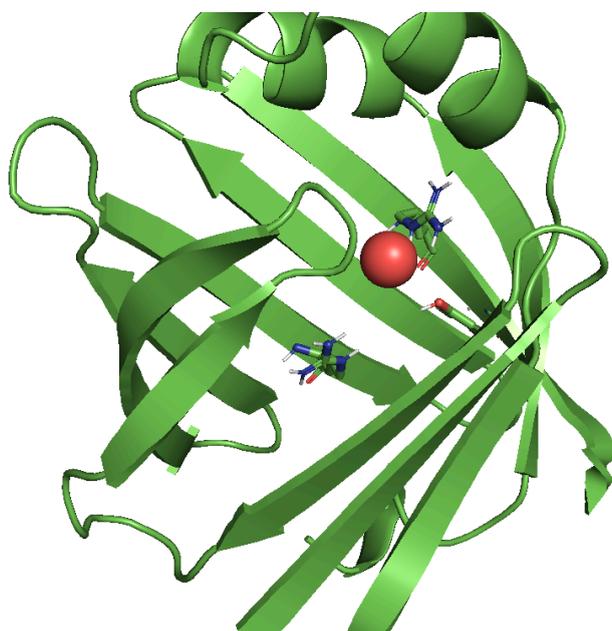


Figura 4.9b. Modelo de *docking* **-D-** utilizado para el **CV** prospectivo. El sesgo donador es indicado por la esfera roja entre los residuos Arg127 y Tyr129.

4.10 SELECCIÓN DEL VALOR DE CORTE PARA EL CRIBADO

De la misma manera que fue mencionado anteriormente en este capítulo, se analizaron las superficies de **PPV**, para las quimiotecas simuladas **LUDe** y **DUDE-Z**, buscando nuevamente la optimización de la **Sp** por sobre la **Se**.

El valor de corte seleccionado para la campaña de **CV** por *docking* fue de -12,65, es decir, compuestos con un *score* igual o más negativo son considerados como **ACTIVOS**, caso contrario se considerarán **INACTIVOS**. Para dicho valor de corte, tenemos un valor de 0,999 para **Sp**, de 0,043 para **Se** y de 0,292 para **PPV** (**Ya**=0,01) en la quimioteca **LUDe**.

Para el caso de **DUDE-Z**, los valores de **Sp**, **Se** y de **PPV** ($Y_a=0,01$) fueron 0,999; 0,043; y 0,709, respectivamente para el valor de corte seleccionado de -12,65. Esta quimioteca presenta un valor máximo de **PPV** de 0,802, el cual se corresponde con un punto de corte de -12,28. Para este último score, el valor de **PPV** ($Y_a=0,01$) para **LUDe** es de 0,192. Por este motivo, se optó por mantener el score de corte en -12,65 para que el valor de **PPV** ($Y_a=0,01$) se encuentre entre aproximadamente 0,3 y 0,7 en lugar de entre 0,2 y 0,8. De esta manera, la probabilidad de encontrar un compuesto activo experimental seleccionado por el modelo oscila entre 3 a 7 de cada 10, considerando un 1% de activos. Las **Figuras 4.10a,b** y la **Tabla 4.10** muestran los resultados.

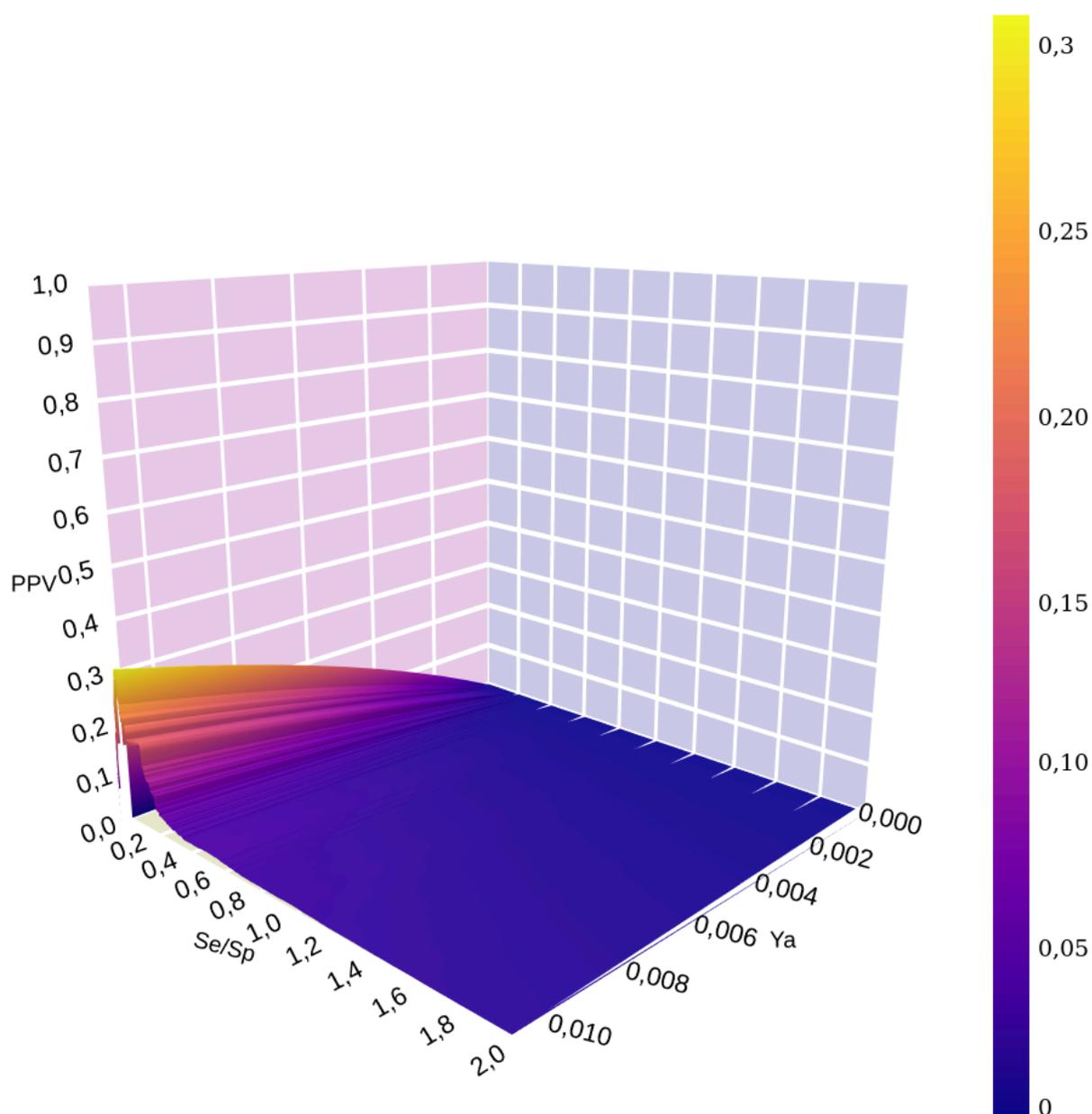


Figura 4.10a. Superficie de **PPV** para el modelo de *docking -D-* con la quimioteca **LUDe**.

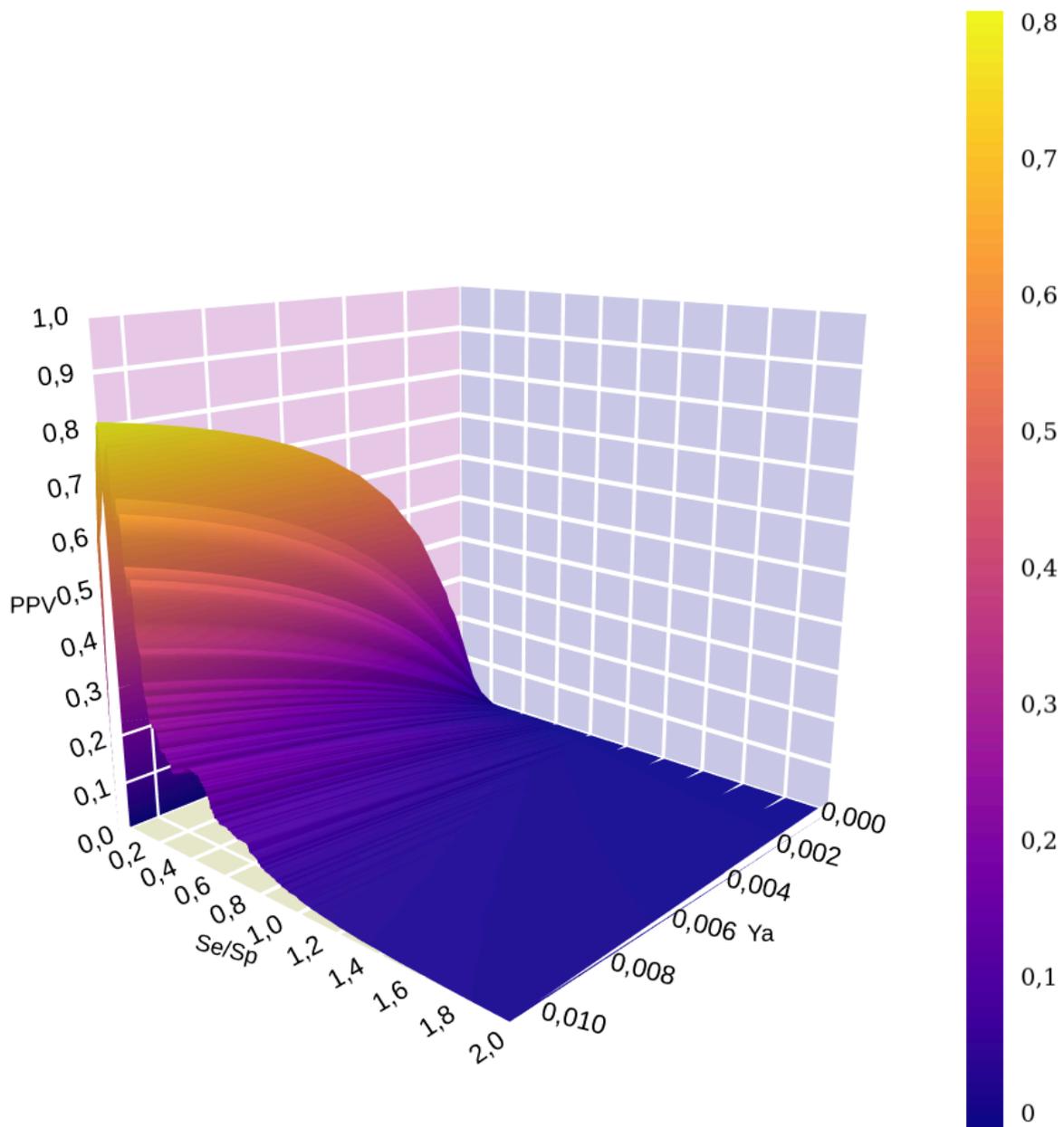


Figura 4.10b. Superficie de **PPV** para el modelo de *docking -D-* con la quimioteca **DUDE-Z**.

Tabla 4.10. Valores de **Se**, **Sp** y **PPV** para los valores de corte analizados hacia el **CV** prospectivo. Los valores entre paréntesis refieren a la proporción de compuestos activos considerada para el cálculo de **PPV**. Los valores en negrita se corresponden con el valor de corte elegido.

QUIMIOTECA	VALOR DE CORTE	Se	Sp	PPV (0,001)	PPV (0,005)	PPV (0,01)
LUDe	-12,65	0,043	0,999	0,039	0,170	0,292
	-12,28	0,074	0,996	0,023	0,106	0,192
DUDE-Z	-12,28	0,074	0,999	0,287	0,669	0,802
	-12,65	0,043	0,999	0,194	0,548	0,709

4.11 CRIBADO VIRTUAL

De la misma manera que se llevó adelante con métodos indirectos, las bases de datos DrugBank, DRH, Sweetlead, HybridMolDB, NuBBEDB, NaturAr, COCONUT y LIDeB fueron cribadas utilizando el modelo de *docking -D-*, a partir de la estructura cristalina de **EgFABP1**. Como resultado, se obtuvieron 98 compuestos que fueron identificados como **ACTIVOS**.

Dentro de esos compuestos activos, 2 de ellos se encuentran en algún estadio de ensayos clínicos para otras patologías y 14 de ellos provienen de la base de datos COCONUT, la cual compila compuestos naturales. En función de su accesibilidad en términos de costo y disponibilidad para su adquisición, se seleccionó un solo compuesto para su evaluación experimental, tratándose del antialérgico Montelukast, apreciándose su estructura química en la **Figura 4.11a**.

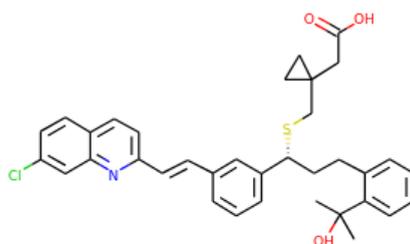


Figura 4.11a. Estructura química del Montelukast, compuesto clasificado como **ACTIVO** por el modelo de *docking*.

En tren de estudiar las diferencias entre las distintas isoformas de las **FABPs**, los 98 compuestos **ACTIVOS** por **EgFABP1** se dockearon utilizando las diferentes conformaciones generadas por **DM** y detalladas en la sección 4.7 del presente capítulo. Para ello, todas las conformaciones fueron alineadas estructuralmente usando como referencia a **EgFABP1** y se configuró el sesgo donar en la misma región del espacio definido para el cristal. En la **Tabla 4.11b** se muestran la cantidad de *hits* por isoforma dentro de esos 98 compuestos **ACTIVOS** por **EgFABP1**.

Tabla 4.11b. Número de compuestos químicos seleccionados en el modelo de **EgFABP1** y en las distintas isoformas. El * indica que el Montelukast se encuentra dentro de los *hits* en la isoforma estudiada mientras que el # indica que esa isoforma está disponible para la evaluación experimental.

HITS POR ISOFORMA	EgFABP1.2	EgFABP2	EgFABP3	EgFABP5
	4	3	7	0
HITS POR ISOFORMA	EmFABP1 #	EmFABP2	EmFABP3 #	EmFABP5
	48	2	35*	1
HITS POR ISOFORMA	TsFABP1	TsFABP2	TsFABP3	TsFABP5
	12	0	29*	12

Debido a la diferente cantidad de *hits* para cada una de las isoformas, se decidió estudiar a las cavidades de los sitios de unión de las distintas conformaciones, para cada una de las isoformas, junto con la estructura cristalina **EgFABP1**. Para ello, como se mencionó en el capítulo 3, se utilizó el servidor DogSiteScorer (Volkamer et al., 2010; Volkamer et al., 2012), obteniendo para cada conformación de cada isoforma los descriptores moleculares correspondientes a cada sitio de unión. Seguidamente, se eliminaron los descriptores cuya varianza fuera menor al 5%, y aquellos que tuvieran una correlación mayor a 0,7, quedando unos 20 descriptores de los 49 iniciales. Sobre estos 20 descriptores se aplicó **PCA**, y posteriormente se normalizaron los datos. Debido a que el primer componente (**PCA1**) describe el 99,77% de la varianza de los datos, se decidió hacer un análisis de *clustering* de los sitios de unión por *k-means*, teniendo en cuenta únicamente a **PCA1** como variable. Como ya se realizó con anterioridad, se evaluó el comportamiento del número de k-grupos en función de la métrica **SIL** para encontrar el número óptimo de *clusters*. Los resultados se muestran en la **Figura 4.11b**, siendo 2 el número óptimo de *clusters* con un **SIL** promedio de 0,628.

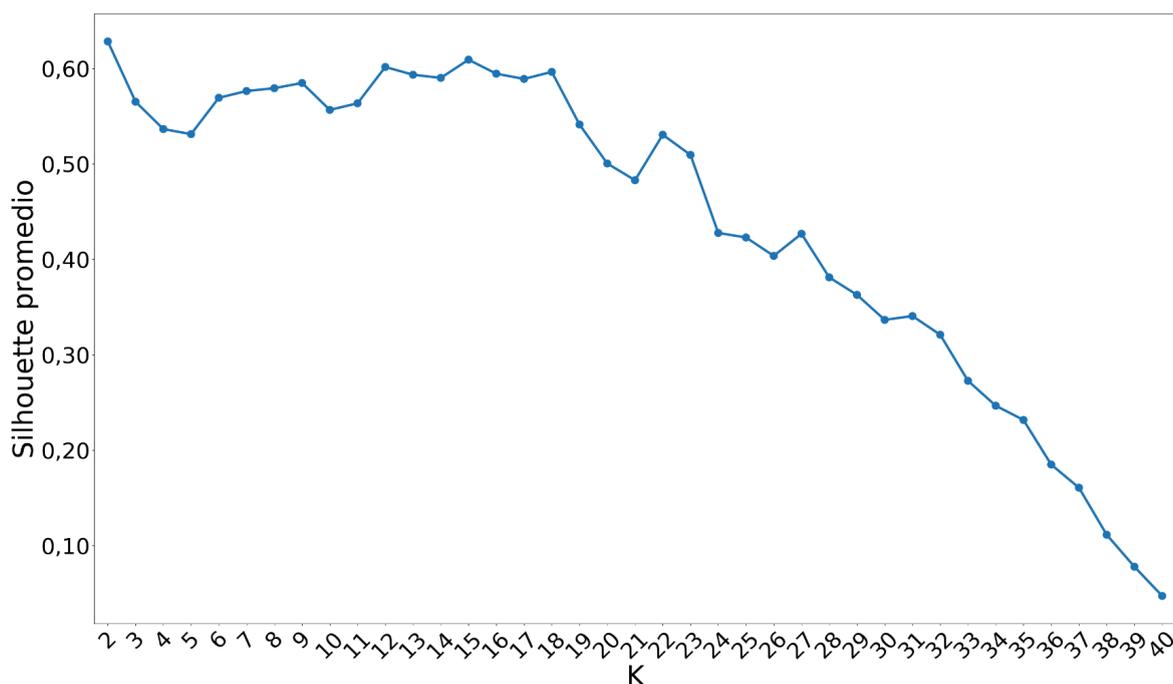


Figura 4.11b. Variación del coeficiente **SIL** en función del valor de **k**.

El sitio de unión de la estructura cristalina de **EgFABP1** forma parte del cluster 1, al igual que dos de las tres poses de **EmFABP1**, tres de las cinco poses de **EmFABP3**, y todas las poses de **TsFABP3**, las tres isoformas que más cantidad de *hits* comparten con **EgFABP1** (Tabla 4.11b). Por otra parte, todas las poses pertenecientes a **EgFABP5**, **EmFABP5** y **TsFABP2** forman parte del cluster 0 siendo estas tres isoformas aquellas con menor cantidad de *hits* en común a **EgFABP1**. Estos resultados parecerían explicar, al menos de manera parcial, la cantidad de compuestos **ACTIVOS** en común entre la estructura de **EgFABP1** y los modelos y la similitud de los sitios de unión. En la **Tabla 4.11c** se muestra la asignación de cada pose para cada isoforma a su respectivo *cluster*.

Tabla 4.11c. Agrupación de la estructura cristalina de **EgFABP1** (PDB:1O8V) y las diferentes isoformas modeladas, según sus características de sitio de unión. Los números seguidos del guión bajo (_) indican la conformación de cada isoforma, cada una de las cuales obtenidas por **DM** y **PCA** en la sección 4.7.

ISOFORMA	CLUSTER	ISOFORMA	CLUSTER
1O8V	1	EmFABP3_0	1
EgFABP1.2_0	1	EmFABP3_1	1
EgFABP1.2_1	0	EmFABP3_2	0
EgFABP1.2_2	1	EmFABP3_3	0
EgFABP1.2_3	1	EmFABP3_4	1
EgFABP2_0	1	EmFABP5_0	0

EgFABP2_1	1	EmFABP5_1	0
EgFABP2_2	1	EmFABP5_2	0
EgFABP3_0	1	TsFABP1_0	1
EgFABP3_1	1	TsFABP1_1	1
EgFABP3_2	1	TsFABP1_2	0
EgFABP3_3	1	TsFABP2_0	0
EgFABP5_0	0	TsFABP2_1	0
EgFABP5_1	0	TsFABP3_0	1
EgFABP5_2	0	TsFABP3_1	1
EmFABP1_0	1	TsFABP5_0	1
EmFABP1_1	1	TsFABP5_1	1
EmFABP1_2	0	TsFABP5_2	1
EmFABP2_0	1	TsFABP5_3	1
EmFABP2_1	1	TsFABP5_4	1
EmFABP2_2	0		

4.12 VALIDACIÓN DE LA POSE EXPERIMENTAL

Según lo descrito en el capítulo 3, se procedió a la validación del *docking power*, es decir, analizar la capacidad de los algoritmos de *docking* ensayados para reproducir la pose experimental de cada uno de los ligandos en cada uno de los receptores estudiados. Para el presente análisis, se utilizaron 33 complejos **FABP**-ligando, en donde se utilizaron **FABPs** de distintas especies y 33 ligandos diferentes, con el fin de que la predicción de la pose pudiera tener cierta generalidad más allá de las isoformas de cestodos.

Los resultados se presentan en las **Figuras 4.12a-e** y la **Tabla 4.12**. Como se puede observar, el algoritmo de **AD Bias** elegido para el **CV** también es el que muestra mejor desempeño entre los algoritmos de *docking* utilizados en términos de **RMSD**. Sin embargo, el **RMSD** promedio para dicho algoritmo es de 3,24 Å, un valor elevado si pensamos en que en los experimentos de *cross-docking* y *re-docking* se suelen considerar adecuados valores inferiores a 3 y 2 Å, respectivamente. No obstante, si separamos del conjunto de datos solamente a las **FABPs** de adipocito y al cristal de **EgFABP1** (16 estructuras en total), el valor de **RMSD** promedio para los experimentos de *cross-docking* desciende a 2,72 Å, logrando el criterio de aceptación recién planteado. Este último valor de **RMSD** es comparable al del cristal

de **EgFABP1**, el cual muestra valores de 2,69 Å para el conjunto de datos completo y de 2,80 Å para el subconjunto de las 16 estructuras.

Por otra parte, la cantidad de re-dockeos exitosos (es decir aquellos cuyo **RMSD** es inferior a 2 Å) fue de 20 en 32 estructuras. Si solo tenemos en cuenta las 16 estructuras recién mencionadas, los re-dockeos exitosos se observaron en nada más 7 de ellas, teniendo un desempeño más desfavorable, comparado con el del conjunto de datos completo, mientras que la *performance* individual para el cristal 1O8V en el *re-docking* fue de 1,22 Å, cumpliendo con el criterio de aceptación planteado.

Por otro lado, si el subconjunto de estructuras es analizado en los experimentos de *cross-docking* de **AD GPU**, el valor de **RMSD** desciende a 3,00 Å (en comparación a los 3,70 Å para el conjunto entero, a los 3,57 Å para 1O8V en el total de estructuras y a los 4,21 Å para 1O8V en el subconjunto de datos). Con respecto a los re-dockeos exitosos, se obtuvieron 7/16 frente al 14/32 del total de los datos, aunque para el cristal de **EgFABP1** el valor de **RMSD** para el *re-docking* fue de 1,22 Å, cumpliendo también con el criterio de aceptación.

Para el caso de **AD Hyd**, el **RMSD** es de 3,76 Å para el *cross-docking* en el subconjunto de las 16 estructuras, contra 4,24 Å teniendo en cuenta el conjunto entero, 3,95 Å para 1O8V en el total de estructuras y 3,91 Å para 1O8V en el subconjunto de datos. Por otro lado, los re-dockeos exitosos fueron 4/16 contra 7/31 en el conjunto total de estructuras, mientras que el **RMSD** de *re-docking* para **EgFABP1** fue de 4,82 Å.

Con respecto a **Vina**, el *cross-docking* en el conjunto reducido arroja un valor promedio de **RMSD** de 4,74 Å contra el valor de 5,25 Å correspondiente con el total de las estructuras, como así también para 1O8V en el conjunto completo o reducido de datos. En cuanto a los re-dockeos exitosos, los mismos fueron de 5/16 contra 9/33 del total de datos, mientras que el **RMSD** de *re-docking* para **EgFABP1** fue de 6,87 Å.

Finalmente, para **Vina + Vinardo**, el *cross-docking* en el conjunto reducido tuvo un valor promedio de **RMSD** de 4,31 Å en comparación al 5,25 Å del total de las estructuras, 5,04 Å para 1O8V en el total de estructuras y 4,10 Å para 1O8V en el subconjunto de datos. En cuanto a los re-dockeos exitosos, **Vina + Vinardo** obtuvo una relación de 6/16 contra 12/33 del total y un valor de **RMSD** para **EgFABP1** de 6,87 Å.

A pesar que **AD Bias** arrojó valores razonables de *cross-docking* para el subconjunto de estructuras mencionadas, la *performance* para el *re-docking* es inferior sobre ese subconjunto con respecto al total de los datos. Por otro lado, los valores de **RMSD** correspondientes al cristal de **EgFABP1** se encuentran en todos

los casos dentro de los criterios de aceptación, situación que no sucede con el resto de los algoritmos. Estos resultados indican que **AD Bias** es capaz de reproducir la pose experimental de distintos ligandos en **EgFABP1**, aunque no es posible afirmar lo mismo para todo el conjunto de datos utilizado.

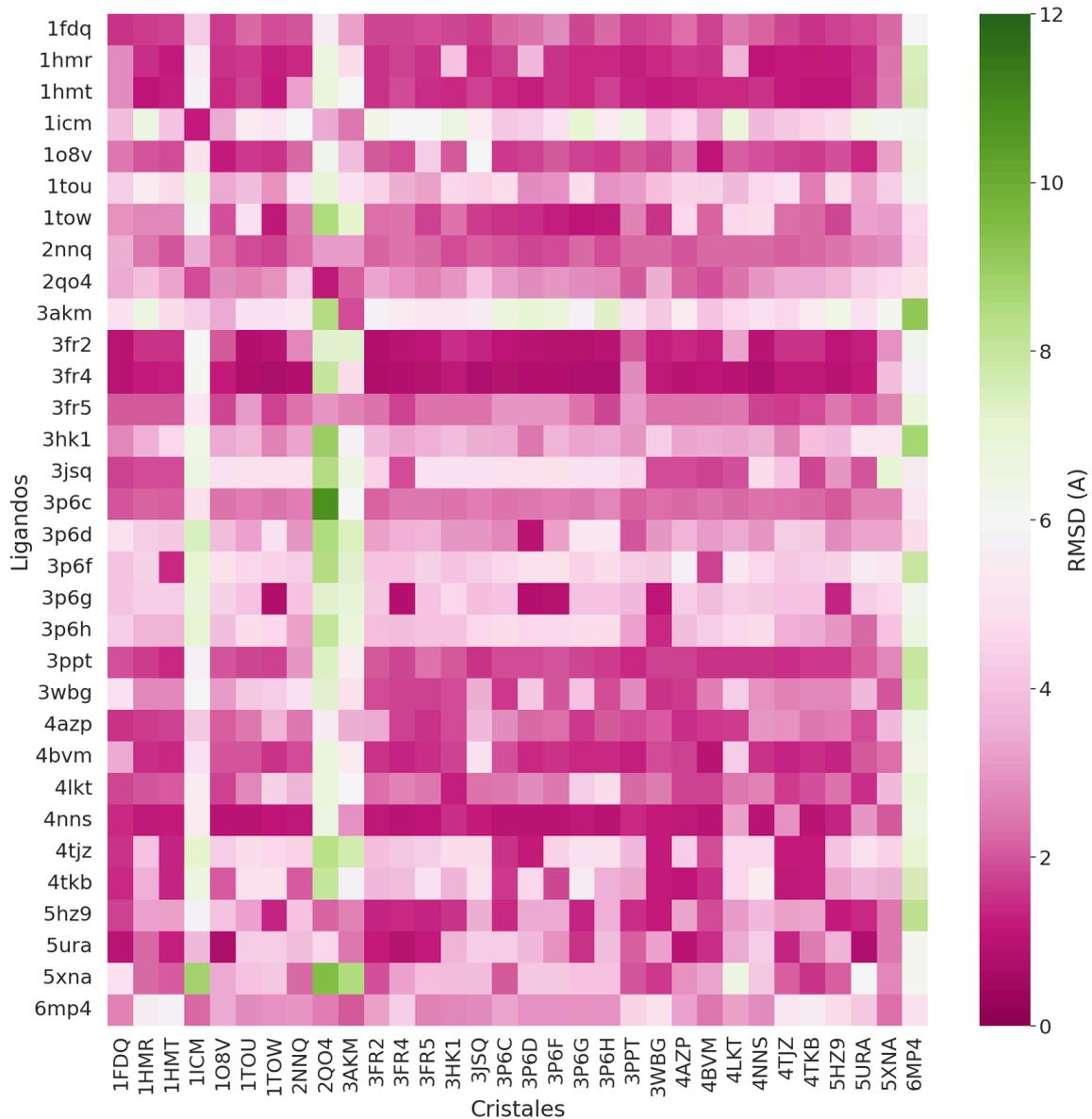


Figura 4.12a. RMSD para cada par cristal-ligando para la mejor pose predicha por **AD Bias** (-D-). Los códigos en mayúsculas se corresponden con el código **PDB** del cristal, mientras que las minúsculas se corresponde con el ligando respectivo.

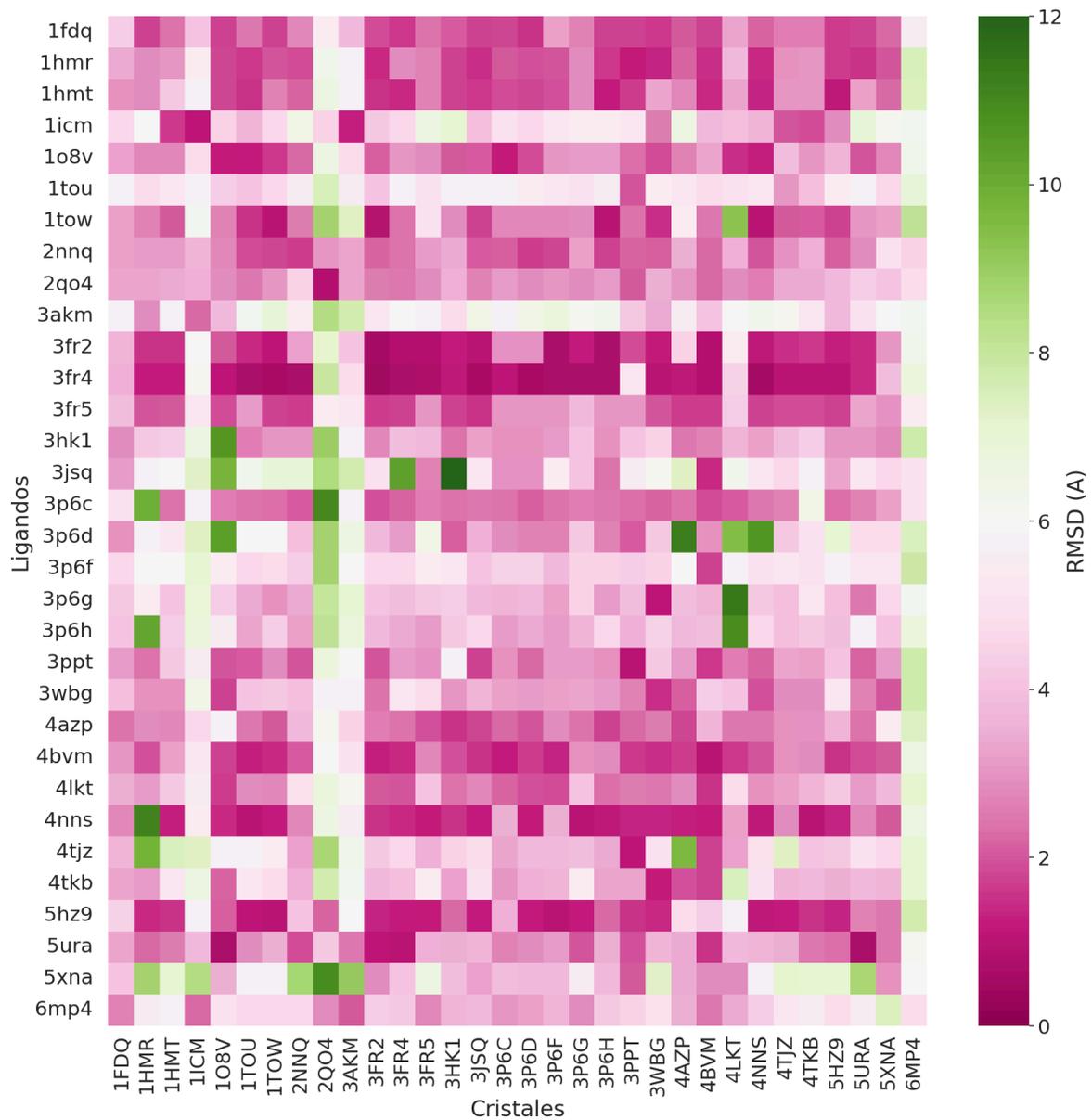


Figura 4.12b. RMSD para cada par cristal-ligando para la mejor pose predicha por **AD GPU**. Los códigos en mayúsculas se corresponden con el código **PDB** del cristal, mientras que las minúsculas se corresponde con el ligando respectivo.

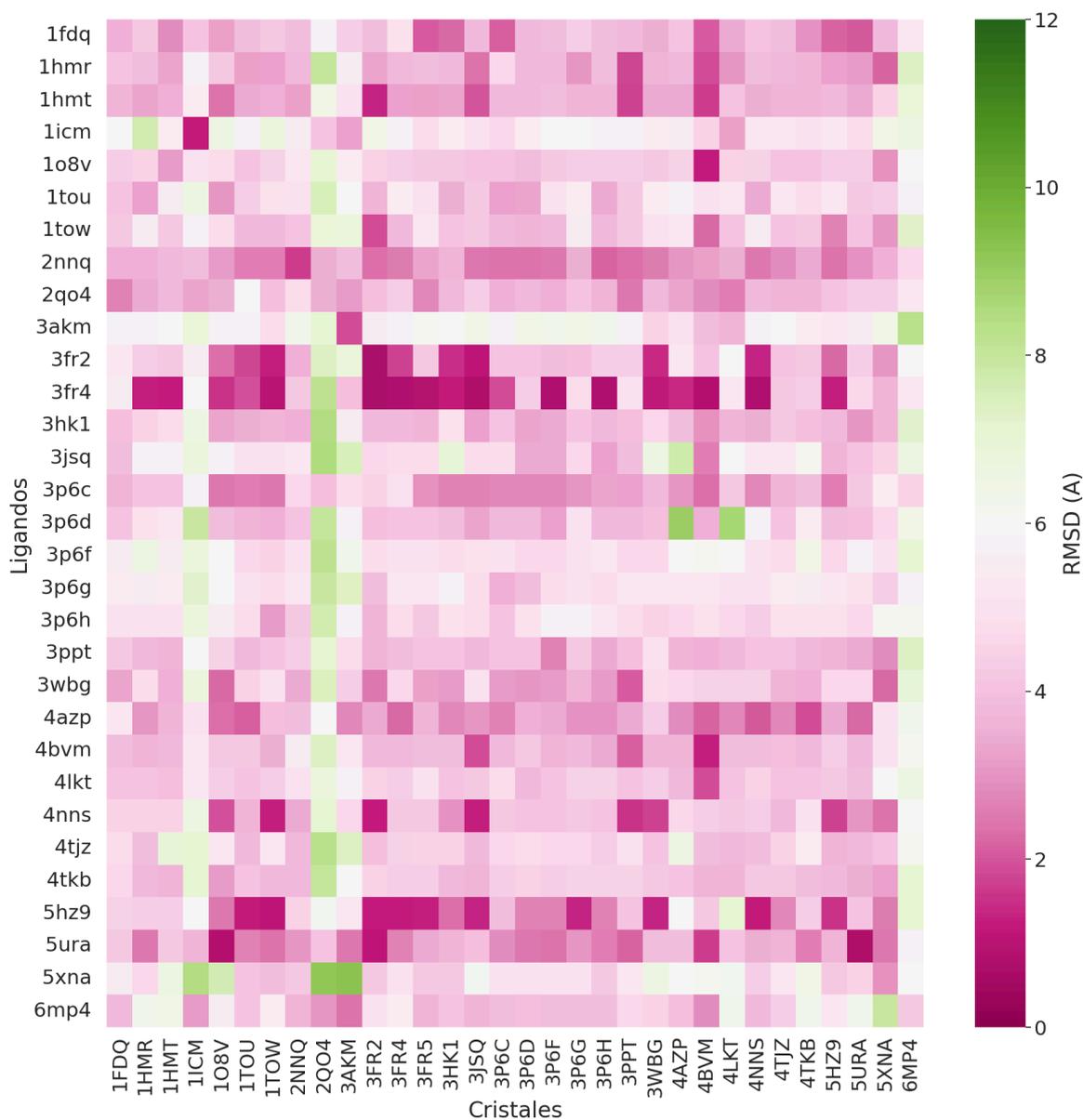


Figura 4.12c. RMSD para cada par cristal-ligando para la mejor pose predicha por **AD Hyd**. Los códigos en mayúsculas se corresponden con el código **PDB** del cristal, mientras que las minúsculas se corresponde con el ligando respectivo.

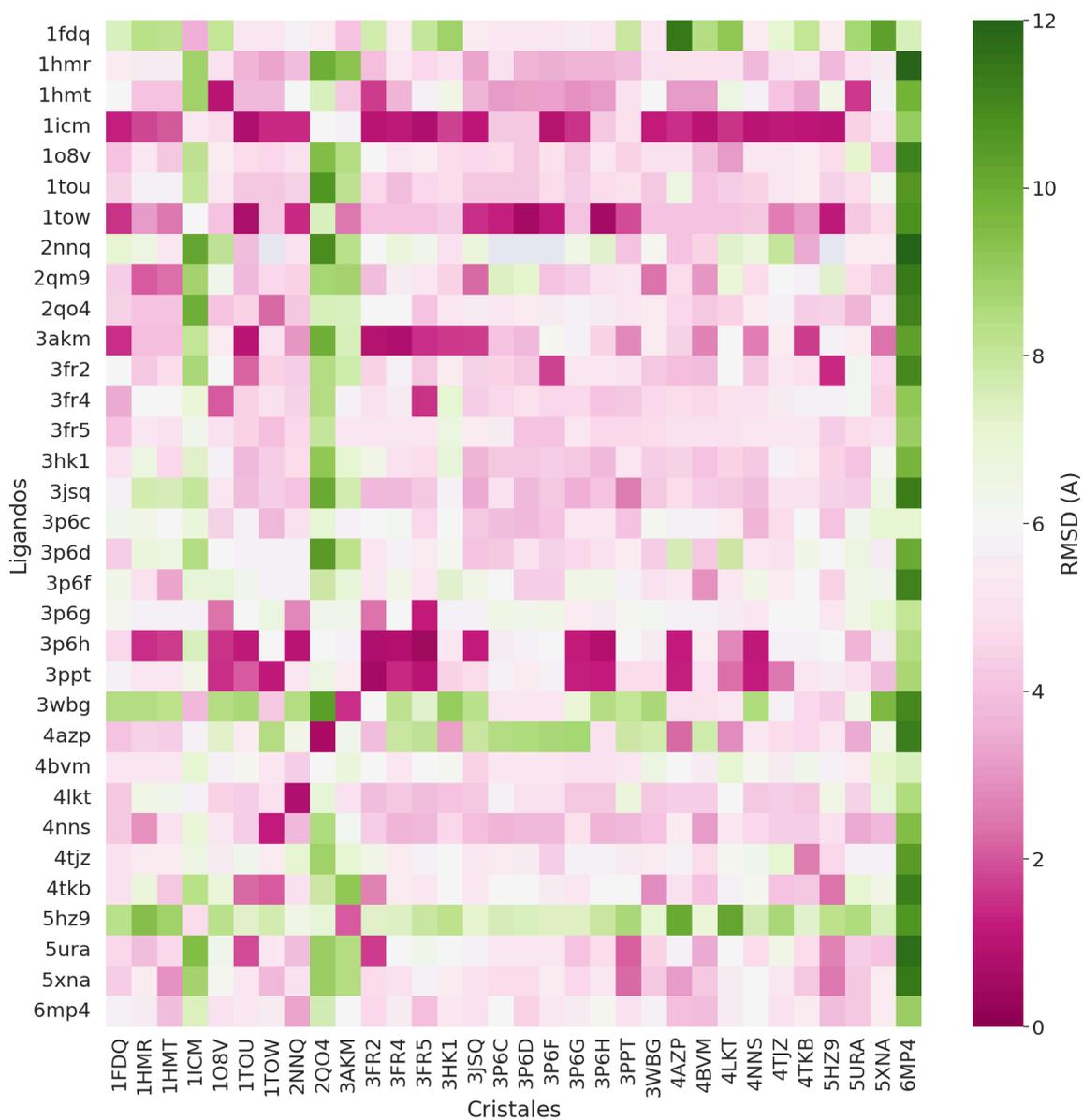


Figura 4.12d. RMSD para cada par cristal-ligando para la mejor pose predicha por **Vina**. Los códigos en mayúsculas se corresponden con el código **PDB** del cristal, mientras que las minúsculas se corresponde con el ligando respectivo.

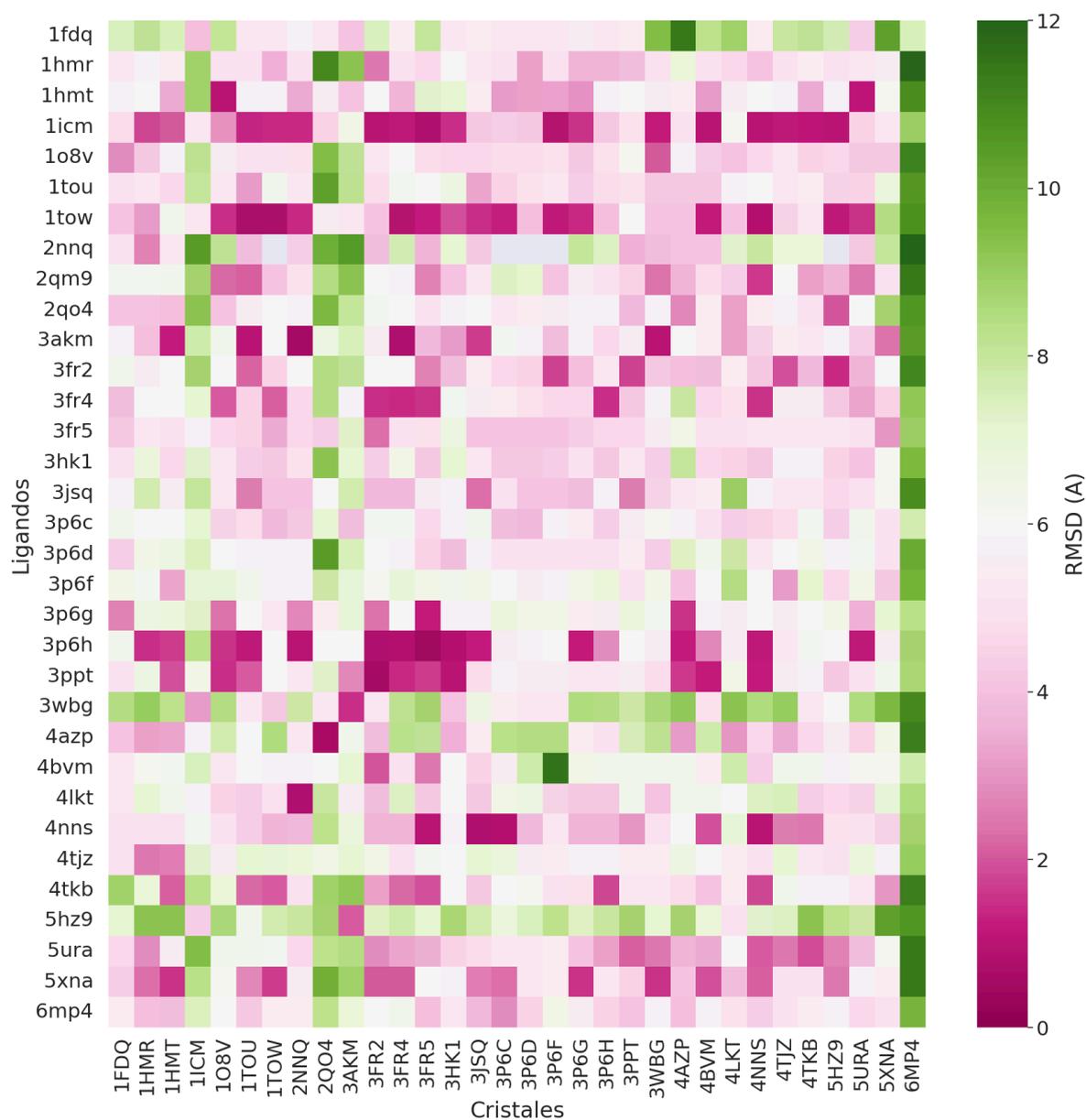


Figura 4.12e. RMSD para cada par cristal-ligando para la mejor pose predicha por **Vina + Vinardo**. Los códigos en mayúsculas se corresponden con el código **PDB** del cristal, mientras que las minúsculas se corresponde con el ligando respectivo.

Tabla 4.12. RMSD promedio para cada uno de los algoritmos de *docking* ensayados y número de *re-dockings* con un **RMSD** menor a 2Å. Los encabezados subrayados indican al subconjunto de estructuras, mientras que los encabezados sin subrayar aluden al conjunto completo de estructuras. Los valores entre paréntesis indican el desvío estándar.

ALGORITMO	RMSD (Å)	<u>RMSD</u> (Å)	RMSD (Å) 108V	<u>RMSD</u> (Å) 108V	RE-DOCKING < 2Å	<u>RE-DOCKING</u> < 2Å
AD Bias (-D-)	3,24 (1,43)	2,72 (0,28)	2,69 (2,19)	2,80 (1,33)	20/32	7/16
AD GPU	3,70 (1,69)	3,00 (0,44)	3,57 (2,62)	4,21 (3,29)	14/32	7/16
AD Hyd	4,24 (1,23)	3,76 (0,37)	3,95 (1,63)	3,91 (1,49)	7/31	4/16
Vina	5,25 (1,70)	4,74 (2,19)	5,25 (1,92)	5,25 (1,92)	9/33	5/16
Vina + Vinardo	5,25 (1,84)	4,31 (0,49)	5,04 (2,15)	4,10 (1,93)	12/33	6/16

4.13 ENSAYOS DE DESPLAZAMIENTO MONITOREADOS POR FLUORESCENCIA

Se ensayaron los compuestos candidatos a unir **FABPs** provenientes tanto de métodos indirectos (Hidroclorotiazida, Fenticonazol y Naratriptán) como de métodos directos (Montelukast) utilizando la técnica de desplazamiento de la sonda fluorescente **ANS**. Para tales propósitos, se purificaron de manera recombinante las isoformas **EgFABP1**, **EmFABP1** y **EmFABP3**, como se describió en el capítulo de metodología, y se enfrentaron a los compuestos recién mencionados. Como control del ensayo, se utilizó Ácido oleico, un ligando conocido de las **FABPs** (Bélgamo et al., 2020).

Las **Figuras 4.13a-c** muestran los espectros de emisión de fluorescencia para **EgFABP1**, **EmFABP1** y **EmFABP3** en función de agregados sucesivos de Ácido oleico, llegando a las concentraciones finales indicadas en cada leyenda del gráfico respectivo. Se puede apreciar que a medida que aumenta la concentración de Ácido oleico, la fluorescencia del pico del **ANS** (alrededor de 500 nm) disminuye, lo cual evidencia el desplazamiento de éste último por el ligando. Esta situación se repite para los otros dos replicados realizados para cada isoforma.

En la misma sintonía, las **Figuras 4.13d-f** muestran los espectros de emisión de fluorescencia para **EgFABP1**, **EmFABP1** y **EmFABP3** en función de agregados sucesivos de Hidroclorotiazida. De manera similar, se aprecia la disminución en la fluorescencia del pico del **ANS** a medida que aumenta la concentración del compuesto ensayado. Esta situación se repite para los otros dos replicados realizados para cada isoforma.

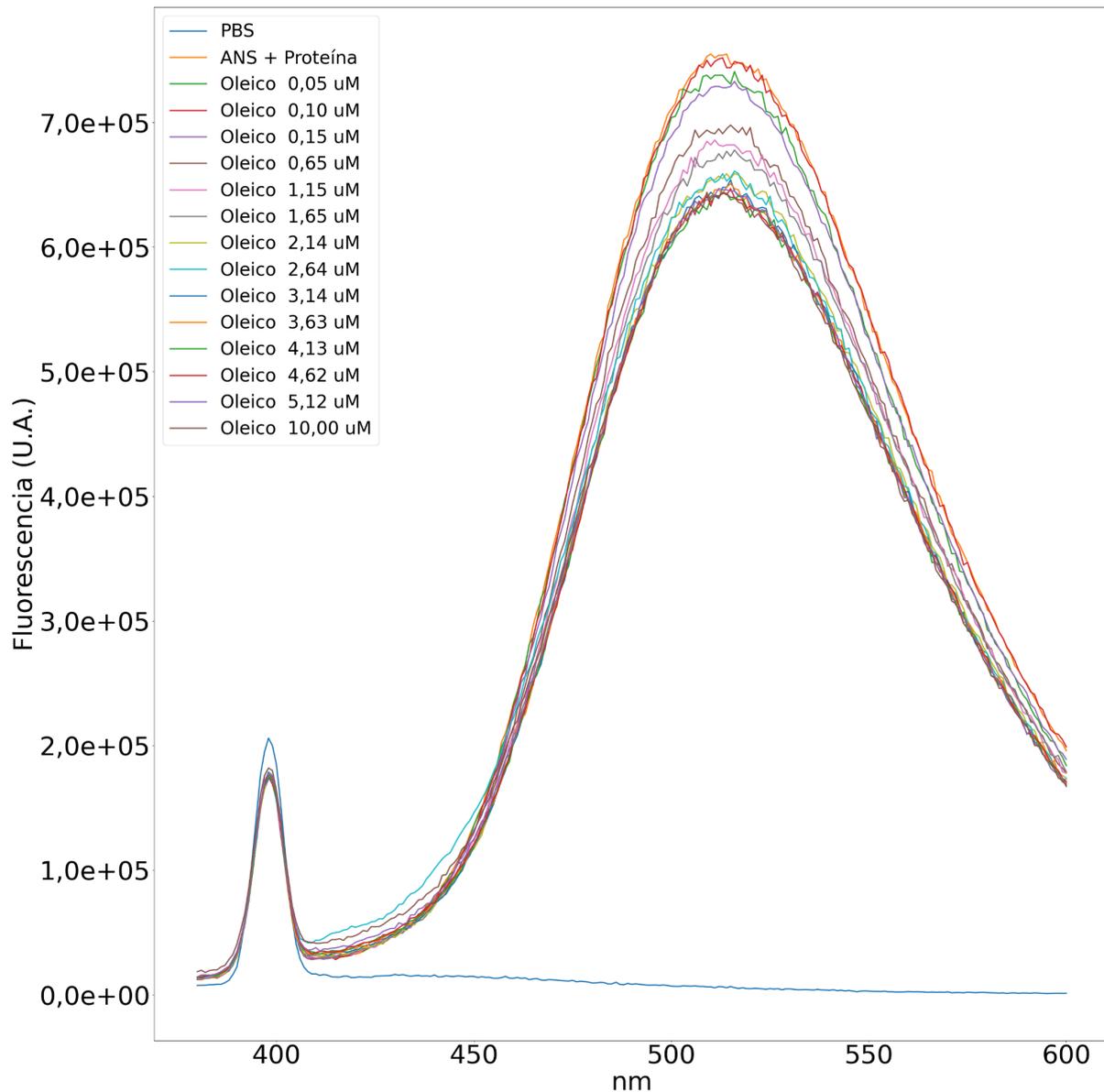


Figura 4.13a. Espectro de desplazamiento de fluorescencia para **EgFABP1** con concentraciones crecientes de Ácido oleico. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM .

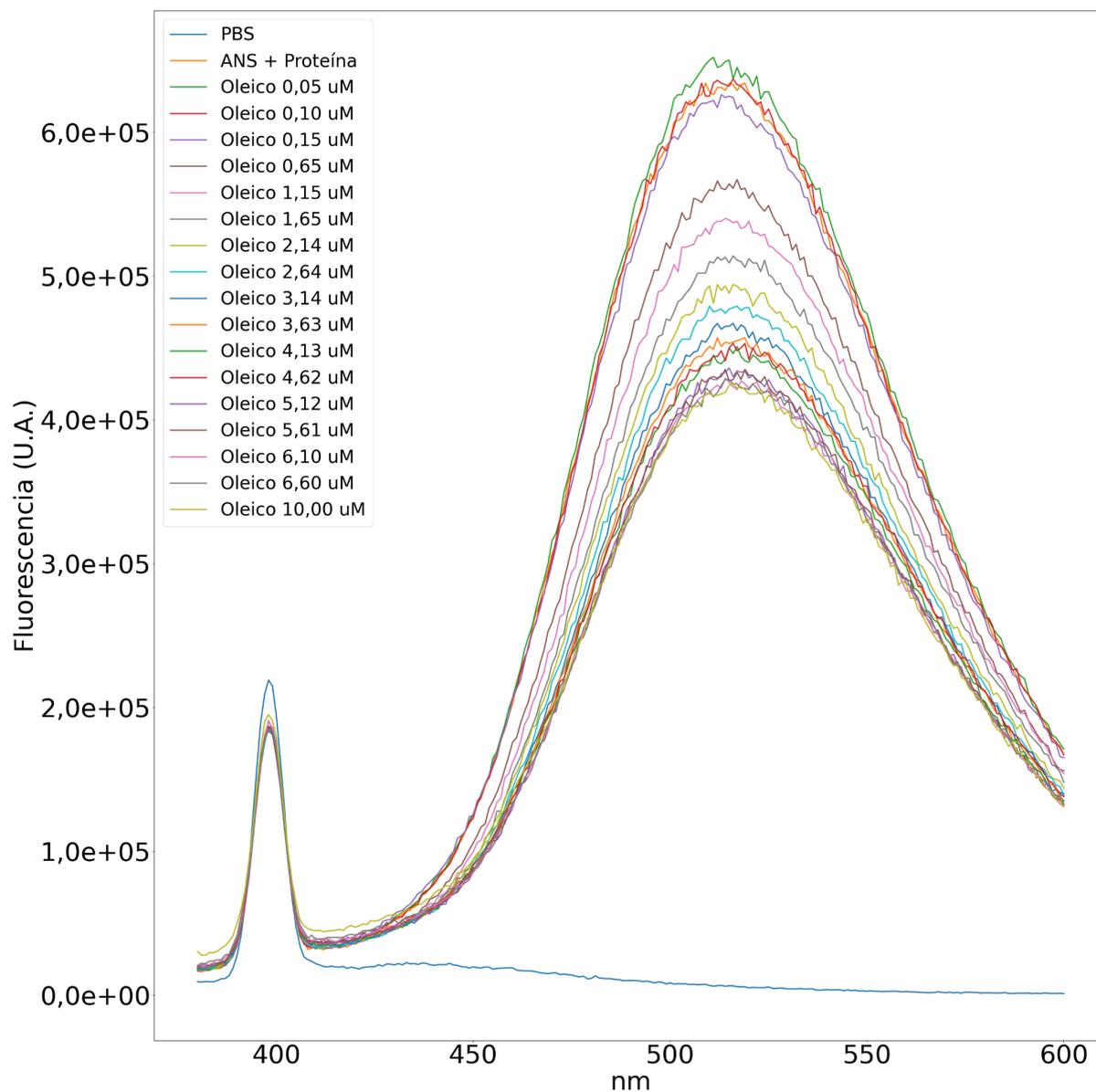


Figura 4.13b. Espectro de desplazamiento de fluorescencia para **EmFABP1** con concentraciones crecientes de Ácido oleico. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM.

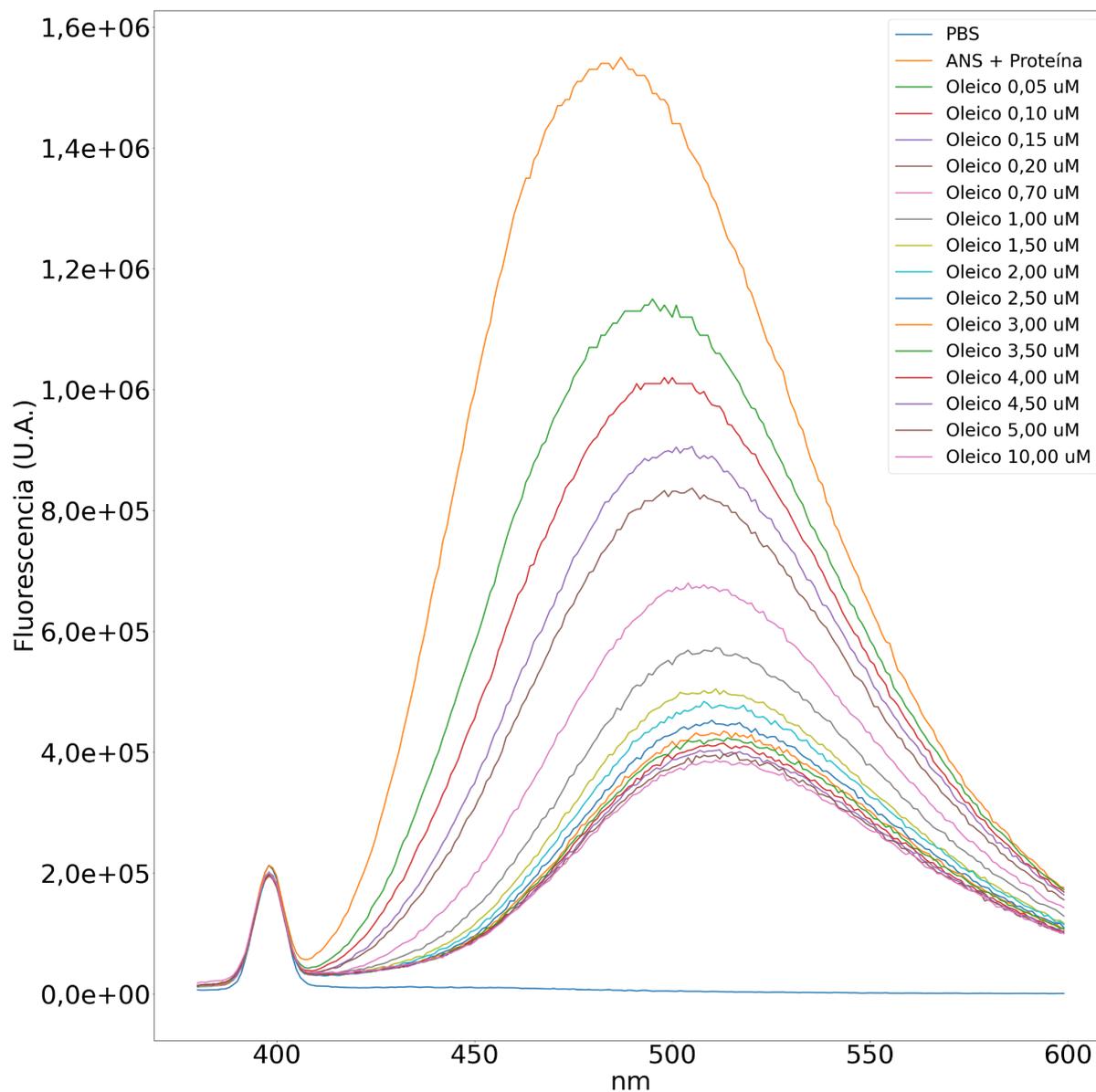


Figura 4.13c. Espectro de desplazamiento de fluorescencia para **EmFABP3** con concentraciones crecientes de Ácido oleico. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM .

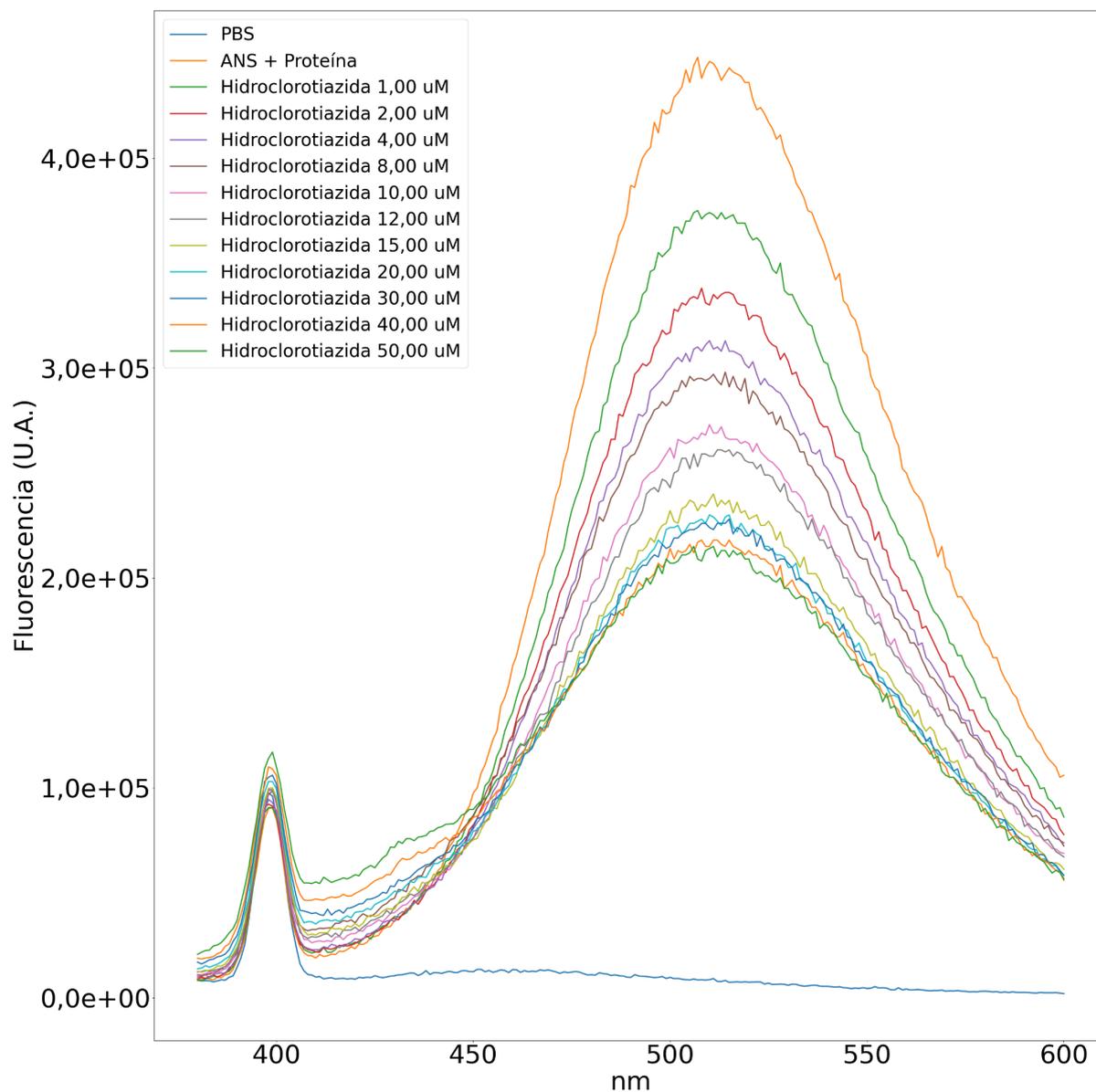


Figura 4.13d. Espectro de desplazamiento de fluorescencia para **EgFABP1** con concentraciones crecientes de Hidroclorotiazida. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM.

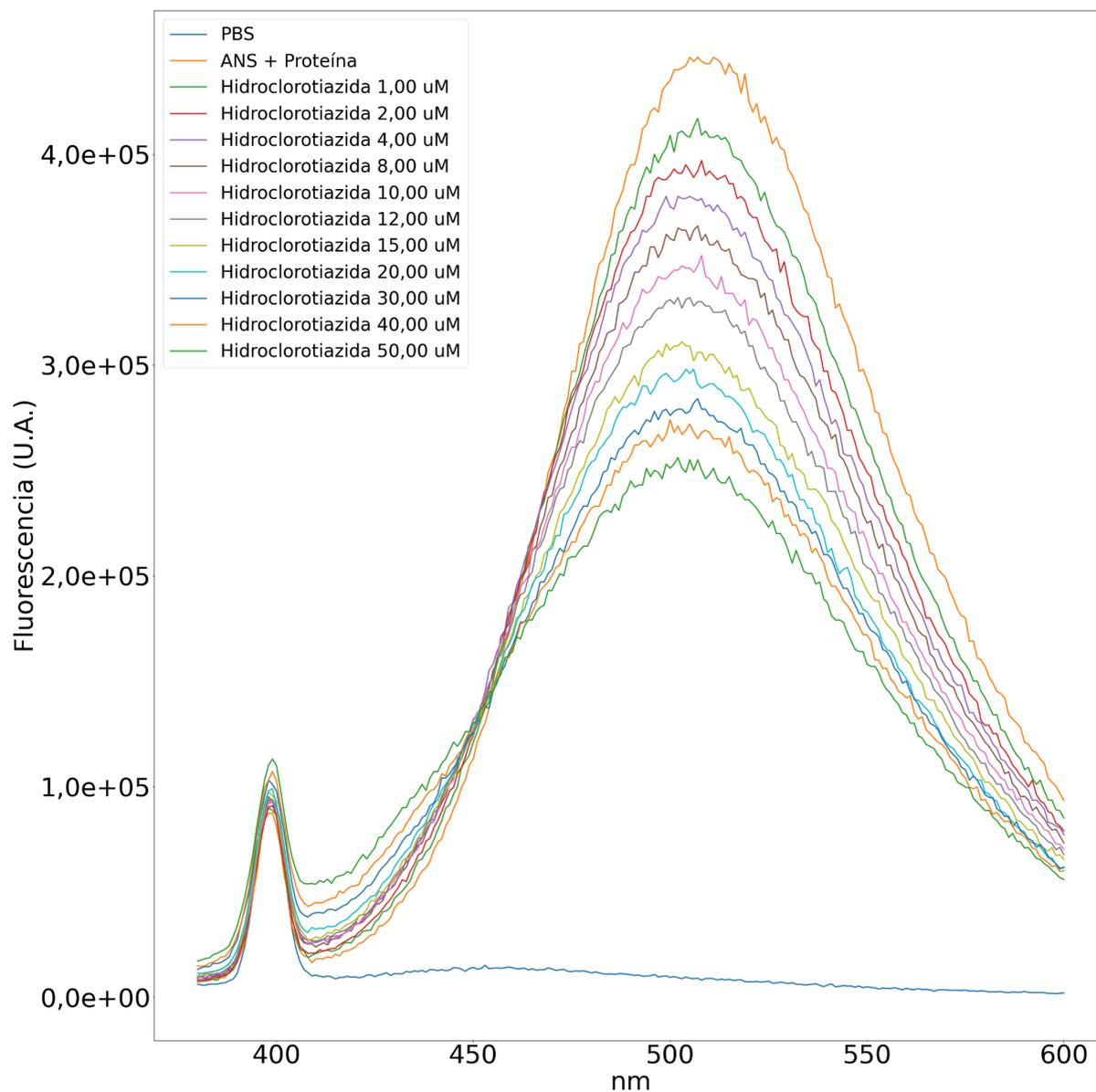


Figura 4.13e. Espectro de desplazamiento de fluorescencia para **EmFABP1** con concentraciones crecientes de Hidroclorotiazida. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM .

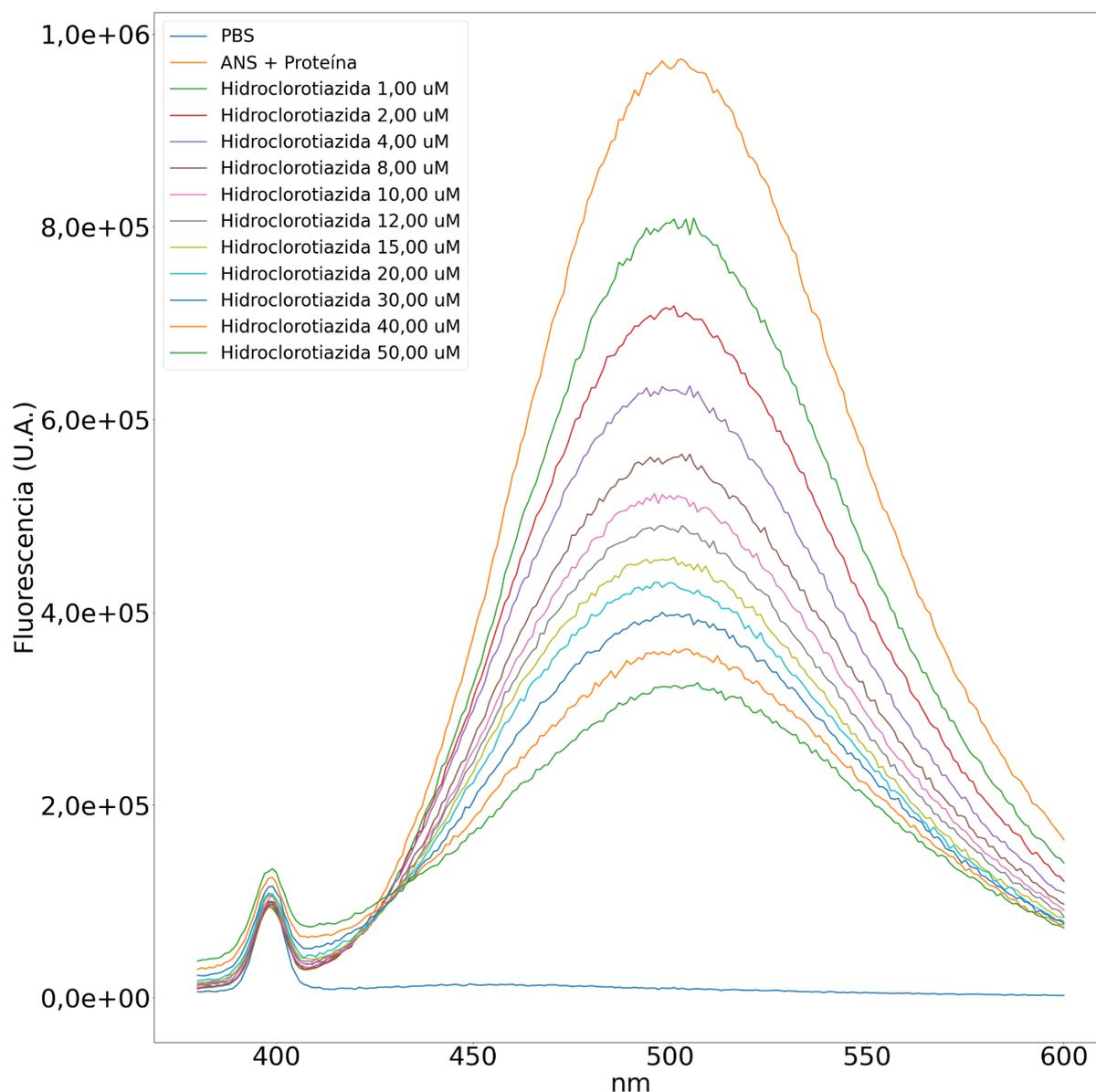


Figura 4.13f. Espectro de desplazamiento de fluorescencia para **EmFABP3** con concentraciones crecientes de Hidroclorotiazida. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM .

Por otra parte, todos los espectros de las **Figuras 4.13a-f** presentan un máximo local alrededor de los 400 nm. Dicho máximo se observa inclusive cuando en la cubeta de medida solo hay buffer **PBS**. Este pico se lo conoce como pico Raman y está relacionado con las vibraciones moleculares del solvente, en este caso del agua (Edinburgh Instruments Ltd., 2023). Muy brevemente, cuando se irradia una muestra con luz, una pequeña fracción de luz dispersada sufre un cambio en su frecuencia debido a su interacción con el solvente. Este cambio en su frecuencia produce que esa fracción de luz pueda interpretarse como luz emitida. La posición de este pico es dependiente de la longitud de onda de excitación que se utilice mediante la **Ecuación 4.2**:

$$1/\lambda_{raman} = 1/\lambda_{exc} - \nu \quad (4.2)$$

donde, λ_{exc} se corresponde con la longitud de onda de excitación del experimento mientras que λ_{raman} se corresponde con la longitud de onda donde se encontrará el pico Raman y ν es el número de onda para el corrimiento de Raman, entre 3400-3600 cm^{-1} en el caso de la molécula de agua (Edinburgh Instruments Ltd., 2023).

Para los fines de este trabajo, el pico Raman no representa un problema, debido a que el mismo se encuentra lejos del máximo de emisión del **ANS**. Sin embargo, para el cálculo de las constantes aparentes de afinidad para cada proteína y cada ligando, se optó por sumar la fluorescencia total para cada punto del espectro entre 420 y 600 nm, dejando así fuera la contribución del pico Raman. Una vez sumada la fluorescencia para cada punto, se normalizaron estos valores para cada punto (donde el valor máximo de fluorescencia se corresponde a concentración cero de compuesto) y se ajustó cada curva de cada replicado al modelo de desplazamiento explicado en el capítulo 3 (**Ecuación 3.10**):

$$\Delta F = \frac{\Delta F_0}{1 + \frac{B}{Kdapp}} + \Delta Fres \quad (3.10)$$

Los resultados obtenidos tanto para el Ácido oleico como para la Hidroclorotiazida se resumen en la **Tabla 4.13**. Se puede observar que la isoforma **EmFABP3** presenta una menor **Kapp** para el Ácido oleico en comparación con las isoformas **EgFABP1** y **EmFABP1**. Por otro lado, **EgFABP1** parece tener mayor afinidad por la Hidroclorotiazida por encima de sus pares **EmFABP1** y **EmFABP3**. Las **Figuras 4.13g-h** muestran una curva representativa de las tres obtenidas para cada par isoforma-compuesto.

Tabla 4.13. Parámetros obtenidos para el Ácido oleico y la Hidroclorotiazida para las tres isoformas en los ensayos de desplazamiento por fluorescencia. Se informa la media de tres replicados junto con el desvío estándar correspondiente.

ISOFORMA	COMPUESTOS	MEDIA DE LOS PARÁMETROS (+/- DE)		
		ΔF_0	K_{app} (μM)	ΔF_{res}
EgFABP1	Ácido oleico	0,18 (0,02)	1,15 (0,43)	0,81 (0,04)
	Hidroclorotiazida	0,39 (0,07)	2,19 (0,51)	0,60 (0,08)
EmFABP1	Ácido oleico	0,42 (0,03)	1,71 (0,29)	0,59 (0,03)
	Hidroclorotiazida	0,42 (0,04)	9,90 (4,33)	0,56 (0,04)
EmFABP3	Ácido oleico	0,65 (0,11)	0,12 (0,06)	0,33 (0,11)
	Hidroclorotiazida	0,42 (0,20)	2,75 (1,64)	0,57 (0,21)

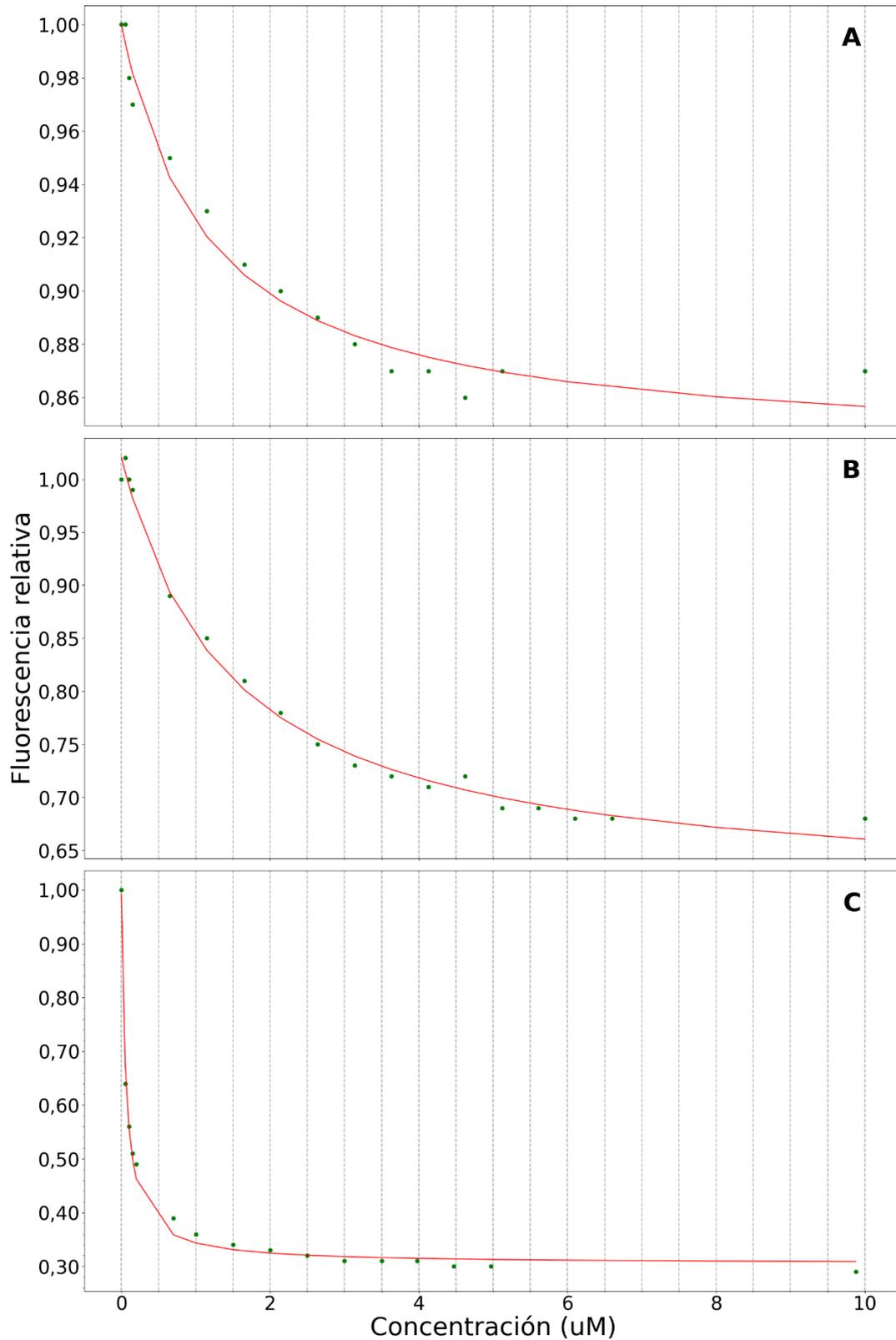


Figura 4.13g. Ajuste de los datos de fluorescencia integrada para cada concentración de Ácido oleico para (A) **EgFABP1**, (B) **EmFABP1**, (C) **EmFABP3**. Cada gráfico es representativo de tres replicados en cada caso.

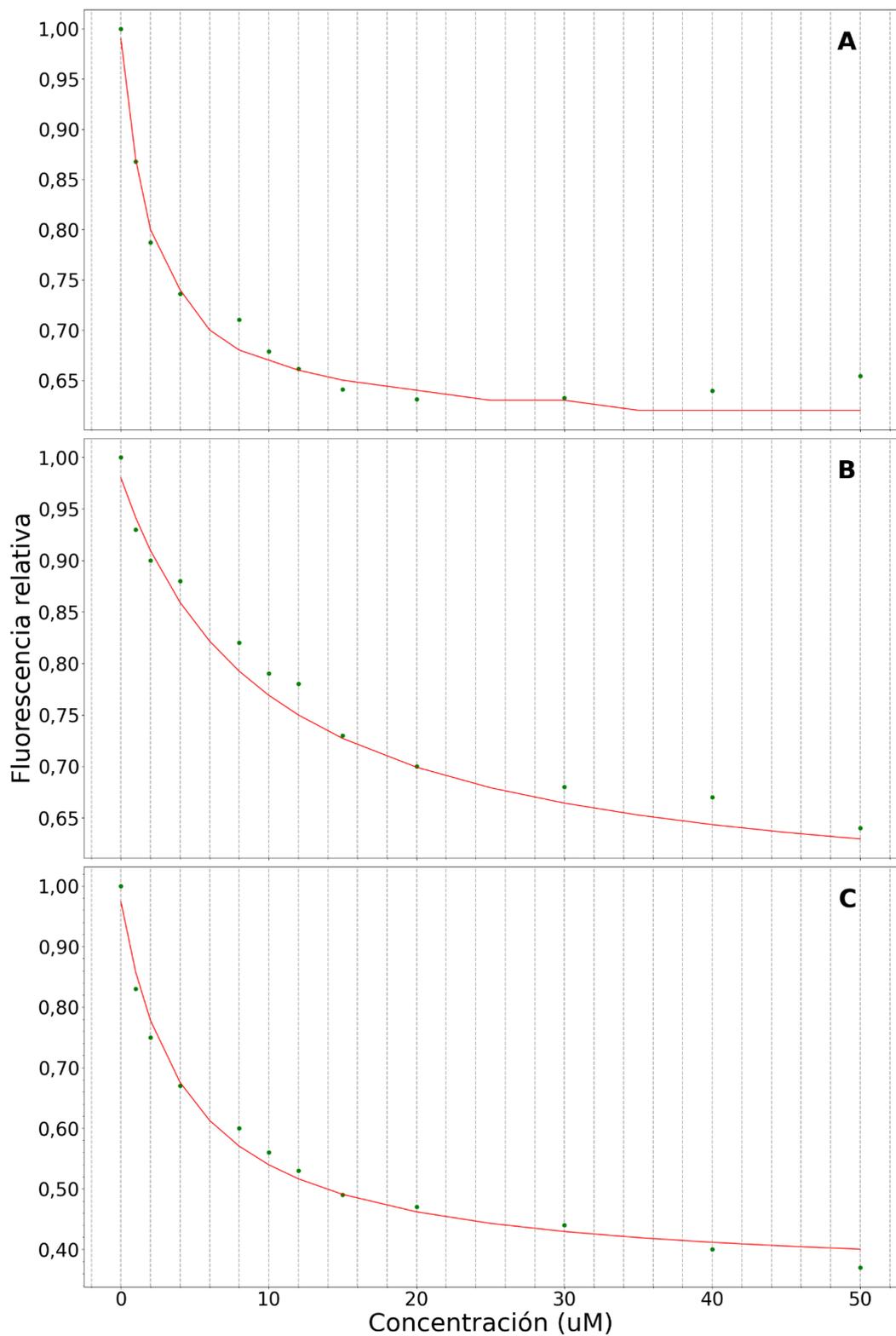


Figura 4.13h. Ajuste de los datos de fluorescencia integrada para cada concentración de Hidroclorotiazida para (A) **EgFABP1**, (B) **EmFABP1**, (C) **EmFABP3**. Cada gráfico es representativo de tres replicados en cada caso.

Por otro lado, se ensayaron los demás compuestos por fluorescencia. En el caso del Montelukast, cuando se llevó adelante el experimento con **EmFABP1**, se observó un aumento en la luz emitida en la región correspondiente al pico Raman, mientras que el máximo correspondiente a la emisión del **ANS** se muestra prácticamente sin cambios (**Figura 4.13i**). Este comportamiento no permite tomar los valores de fluorescencia a partir de 420 nm para el cálculo de la fluorescencia integrada y su posterior relativización a fines del ajuste estadístico, como se hizo con la Hidroclorotiazida y el Ácido oleico.

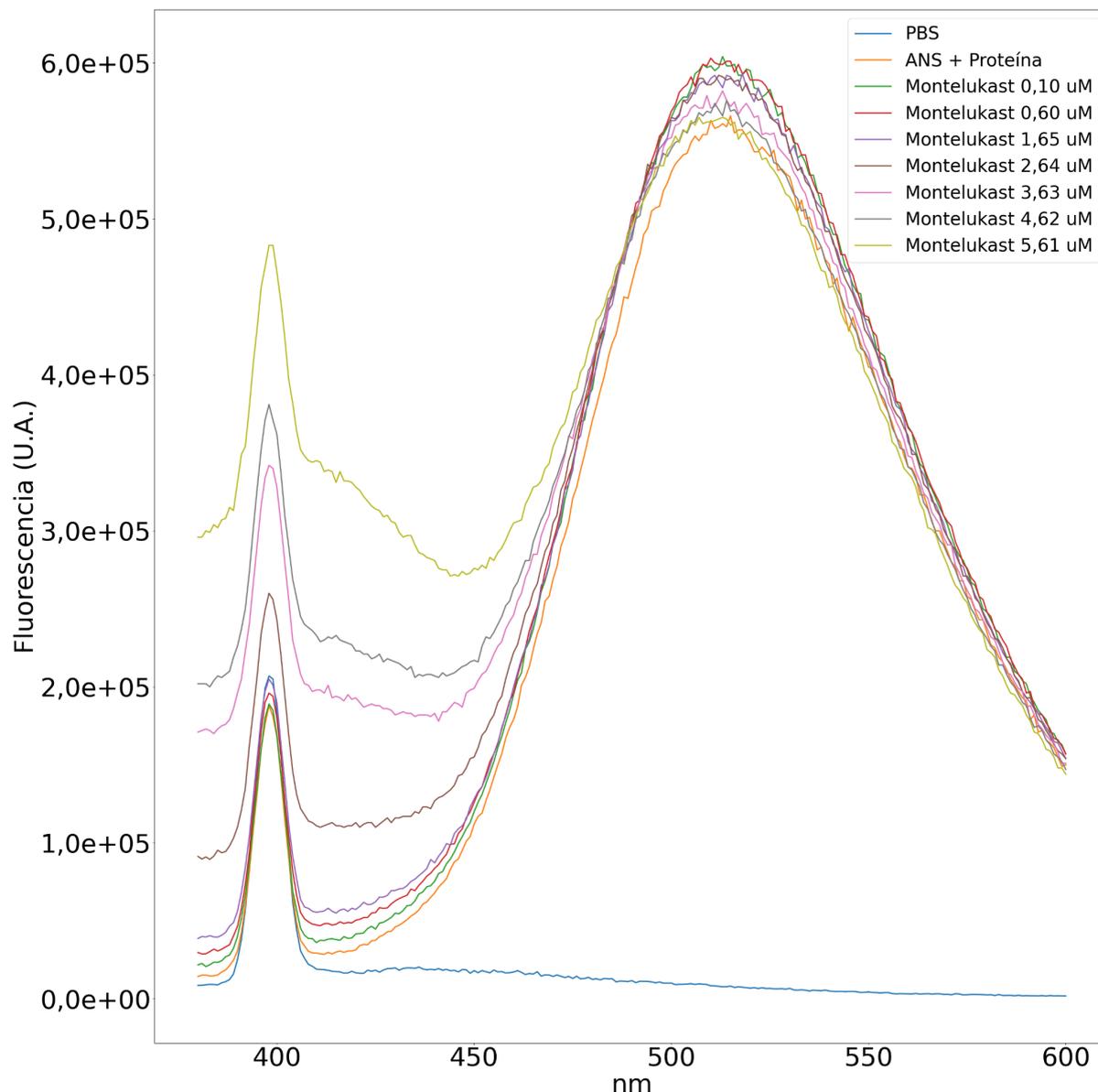


Figura 4.13i. Espectro de desplazamiento de fluorescencia para **EmFABP1** con concentraciones crecientes de Montelukast. La concentración de **ANS** es 20 μM mientras que la de proteína es 1 μM .

Debido a que el Montelukast es fotosensible, se decidió investigar su espectro de absorción (**Figura 4.13j**). En el mismo se puede observar que, en el rango entre

300-600 nm, presenta tres máximos de absorción, correspondientes con 325, 350 y 360 nm aproximadamente. Dado que la sonda **ANS** absorbe también a 350 nm, es una evidencia más a favor de que no es posible medir desplazamiento para este compuesto. Consecuentemente, se decidió llevar adelante experimentos de dicroísmo circular (**CD**) para intentar sortear esta limitación.

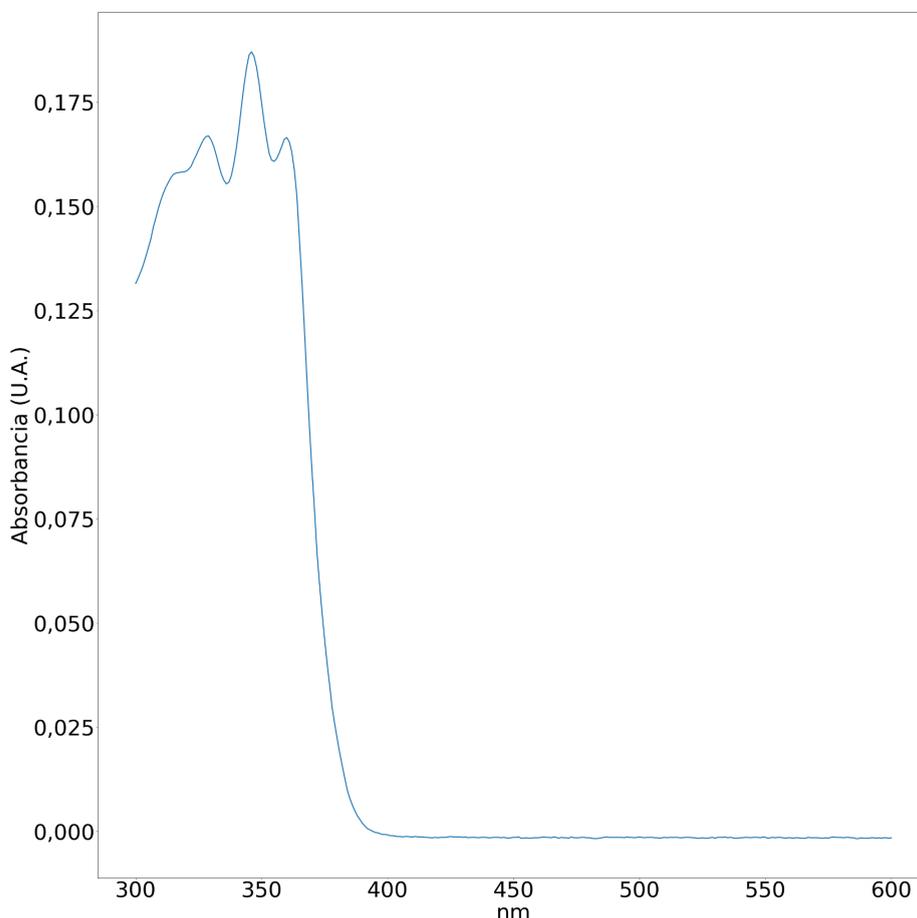


Figura 4.13j. Espectro de absorbancia del Montelukast, a concentración 10 μ M.

Con respecto al Fenticonazol se decidió medir la fluorescencia emitida en presencia de **ANS**, en ausencia y presencia de **EmFABP1** (**Figura 4.13k**). Como se puede observar, la fluorescencia emitida por el **ANS** aumenta con la concentración de Fenticonazol, tanto en ausencia como en presencia de **EmFABP1**. Por lo tanto, este método tampoco pudo ser empleado para la evidenciar unión entre Fenticonazol y **FABPs**.

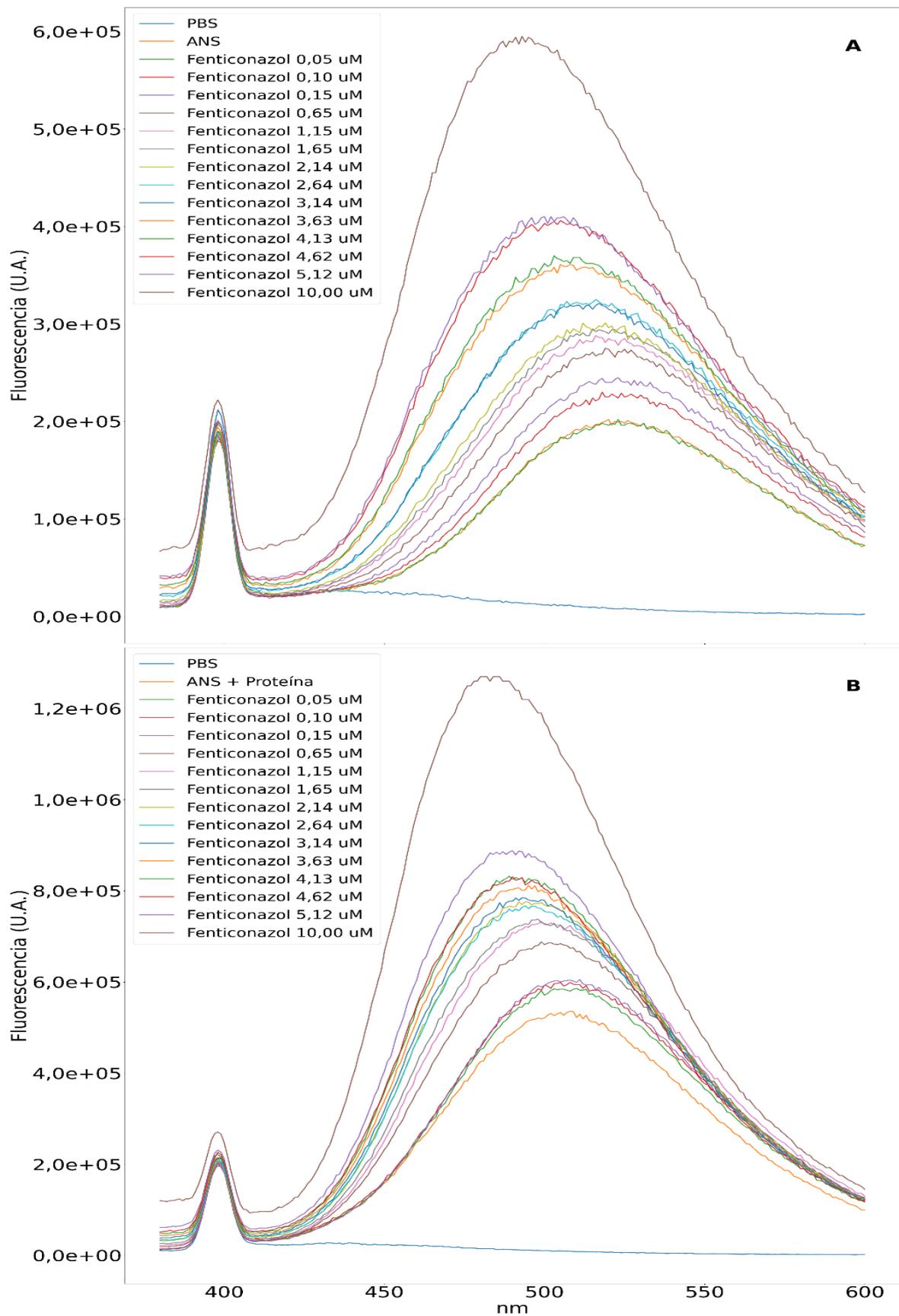


Figura 4.13k. Espectro de emisión de fluorescencia para concentraciones crecientes de Fenticonazol en (A) ausencia y (B) presencia de **EmFABP1**.

Por último, el Naratriptán fue ensayado contra **EmFABP3**. De la misma manera que al Fenticonazol, se lo evaluó en concentraciones crecientes en presencia de **ANS** 20 μM tanto en ausencia como en presencia de proteína. Los resultados se muestran en la **Figura 4.13I**. Como se puede percibir, agregados sucesivos de Naratriptán aumentan los niveles de fluorescencia en todo el espectro de emisión, con lo que el ensayo de desplazamiento no puede ser utilizado para cuantificar **Kdapp** para **FABPs**. Sin embargo, cuando la proteína está presente, se puede observar que hay un corrimiento a la derecha del pico de emisión del **ANS**, lo que podría indicar algún tipo de interacción potencial, aunque sin permitir su cuantificación por éste método.

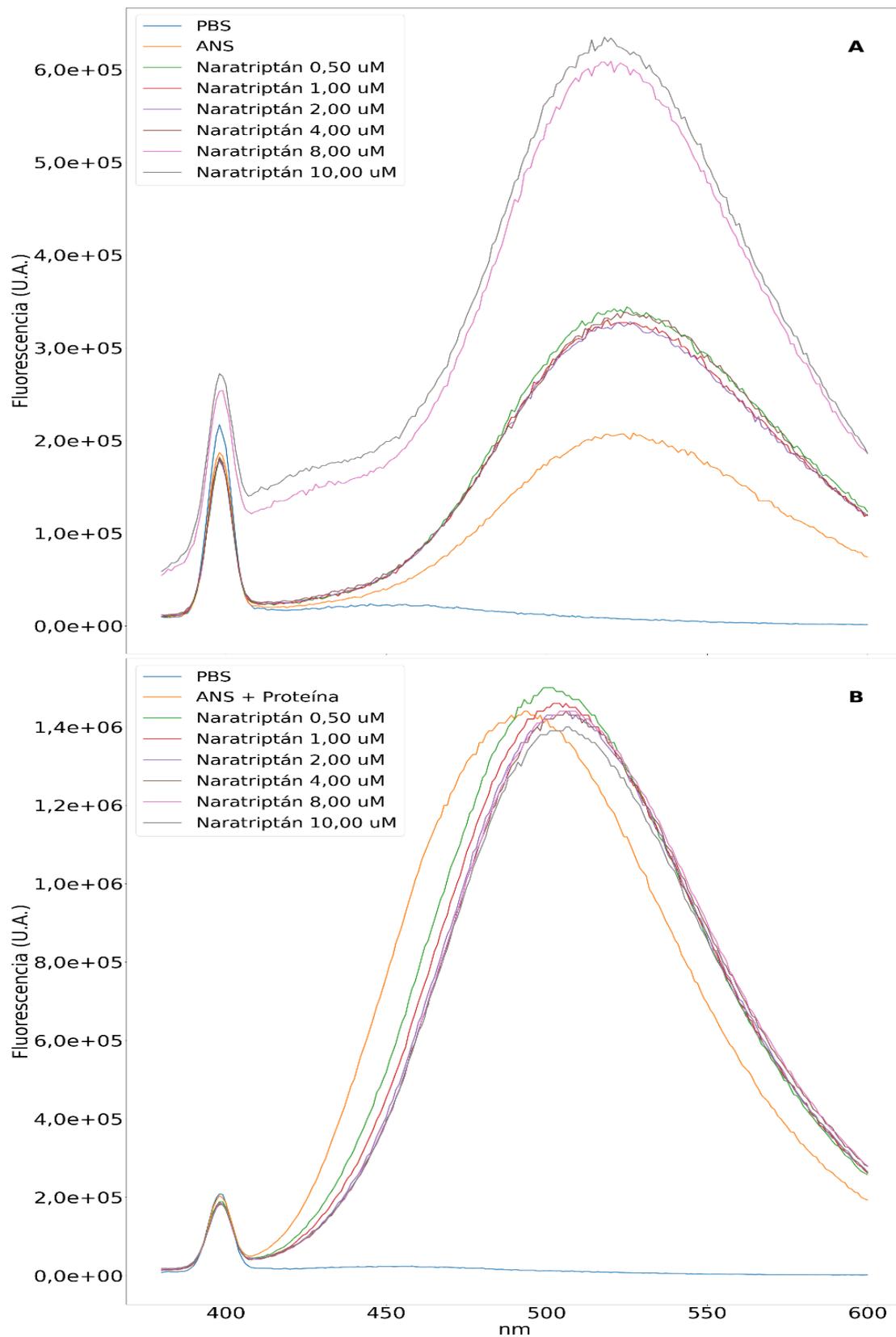


Figura 4.131. Espectro de emisión de fluorescencia para concentraciones crecientes de Naratriptán en **(A)** ausencia y **(B)** presencia de **EmFABP1**.

4.14 ENSAYOS DE DICROÍSMO CIRCULAR

Debido a la imposibilidad de medir desplazamiento de **ANS** por fluorescencia en 3 de los 4 candidatos seleccionados, se decidió enfrentar a **EgFABP1**, **EmFABP1** y **EmFABP3** a diferentes relaciones molares de cada uno de los compuestos y medir en cada caso su espectro de **CD**. En primer lugar, se evaluaron los 4 compuestos candidatos ya mencionados contra las tres isoformas en el *far-UV CD*.

En primer lugar, se observan los espectros del Montelukast contra **EgFABP1**, **EmFABP1** y **EmFABP3** en la **Figura 4.14a**. Como se puede notar, en el caso de **EgFABP1**, a medida que aumenta la relación molar proteína:Montelukast, la fracción del espectro correspondiente a hoja β disminuye (mínimo alrededor de los 220 nm). Esta situación es exactamente inversa para **EmFABP1** y **EmFABP3**. En ninguno de estos casos, el Montelukast mostró absorción de luz circularmente polarizada.

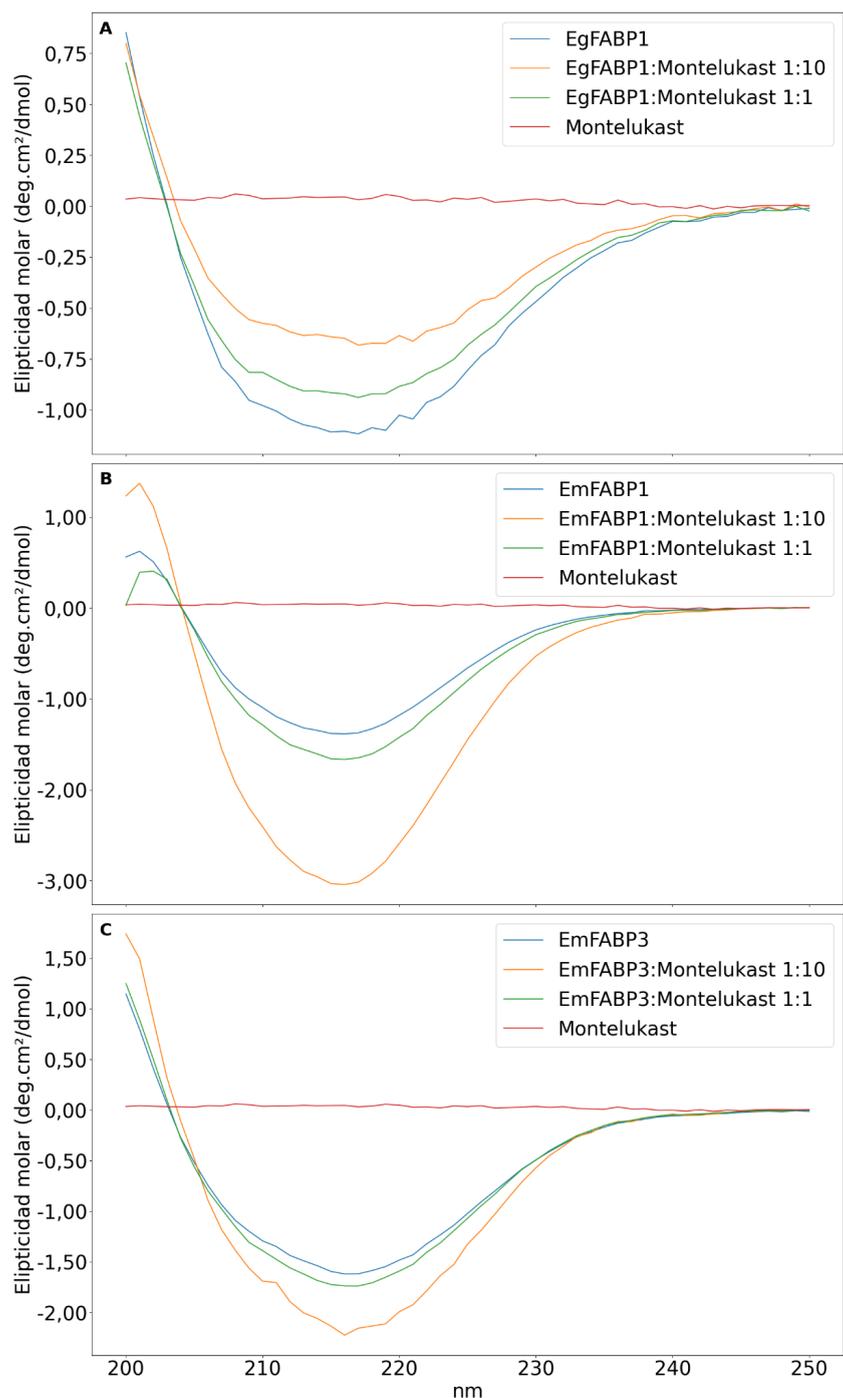


Figura 4.14a. Espectros de *far-UV CD* para Montelukast en presencia de: **(A) EgFABP1** (8 μ M), **(B) EmFABP1** (23 μ M), **(C) EmFABP3** (18 μ M). La concentración de Montelukast en ausencia de proteína fue de 10 μ M.

Seguidamente, se muestran los espectros de la Hidroclorotiazida contra **EgFABP1**, **EmFABP1** y **EmFABP3** en la **Figura 4.14b**. En el caso de **EgFABP1**, la fracción del espectro correspondiente a hoja β disminuye en presencia de Hidroclorotiazida, mientras que para **EmFABP1**, a mayor relación molar proteína:Hidroclorotiazida, la fracción β aumenta. Para **EmFABP3**, pareciera que cuando la relación molar es 1:1

hay una disminución en el contenido de hoja β , situación que cambia cuando la cantidad relativa de Hidroclorotiazida es de 1:10, en la cual hay un aumento en el contenido de dicha fracción, en comparación al espectro de **EmFABP3** en ausencia de compuesto. De la misma manera que con el Montelukast, la Hidroclorotiazida no absorbe luz circularmente polarizada en esta región del espectro UV.

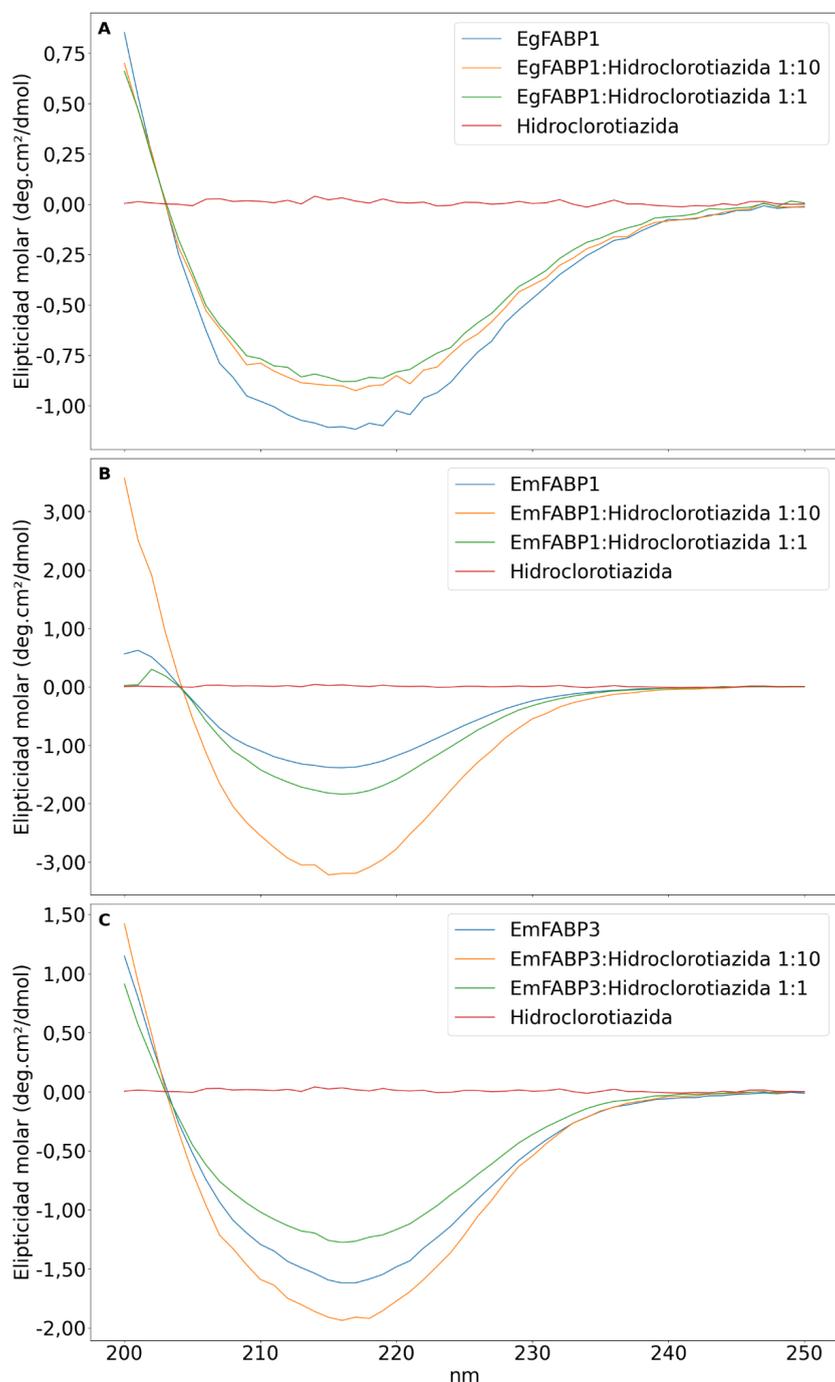


Figura 4.14b. Espectros de *far-UV CD* para Hidroclorotiazida en presencia de: **(A)** **EgFABP1** (8 μM), **(B)** **EmFABP1** (23 μM), **(C)** **EmFABP3** (18 μM). La concentración de Hidroclorotiazida en ausencia de proteína fue de 100 μM .

Por su parte, los espectros del Fenticonazol contra cada una de las isoformas se muestran en la **Figura 4.14c**. Al igual que con la Hidroclorotiazida, la fracción de hoja β es menor en presencia del Fenticonazol para **EgFABP1**. En el caso de **EmFABP3**, se repite también el comportamiento observado para la Hidroclorotiazida. Por otra parte, para **EmFABP1** pareciera haber un aumento en la fracción de hoja β conforme aumenta la concentración del compuesto. El Fenticonazol por sí mismo no es capaz de absorber luz circularmente polarizada.

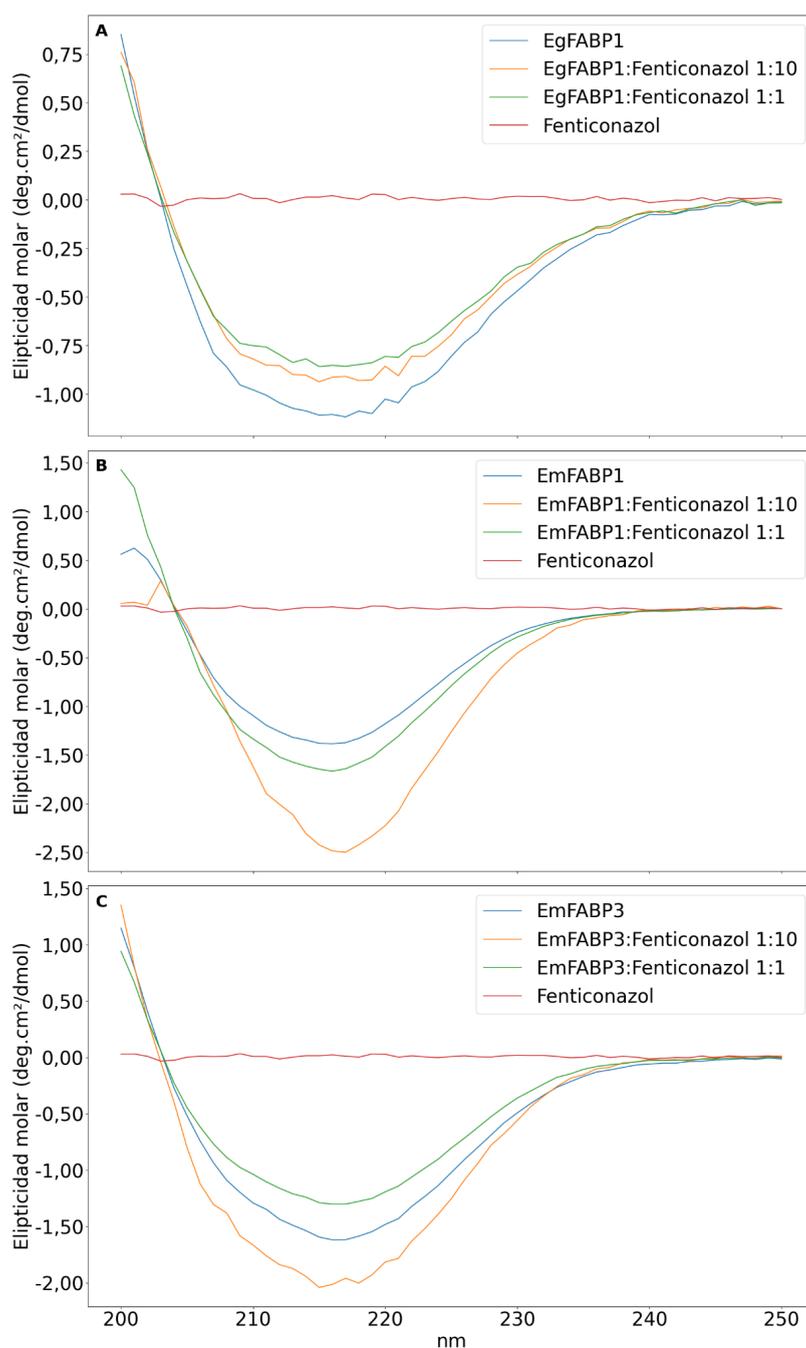


Figura 4.14c. Espectros de *far-UV CD* para Fenticonazol en presencia de: **(A)** **EgFABP1** (8 μM), **(B)** **EmFABP1** (23 μM), **(C)** **EmFABP3** (18 μM). La concentración de Fenticonazol en ausencia de proteína fue de 100 μM .

Finalmente, los espectros del Naratriptán contra **EgFABP1**, **EmFABP1** y **EmFABP3** se muestran en la **Figura 4.14d**. Como se puede notar, en el caso de **EgFABP1**, la fracción del espectro correspondiente a hoja β disminuye en presencia del compuesto (siendo este cambio más pronunciado para la relación 1:1), mientras que para **EmFABP1**, a mayores cantidades de Naratriptán, la tendencia de aumento de la fracción β pareciera ser más pronunciada. En el caso de **EmFABP3** pareciera haber cambios únicamente cuando la relación molar es 1:10. En ninguno de estos casos, el Naratriptán mostró absorción de luz circularmente polarizada.

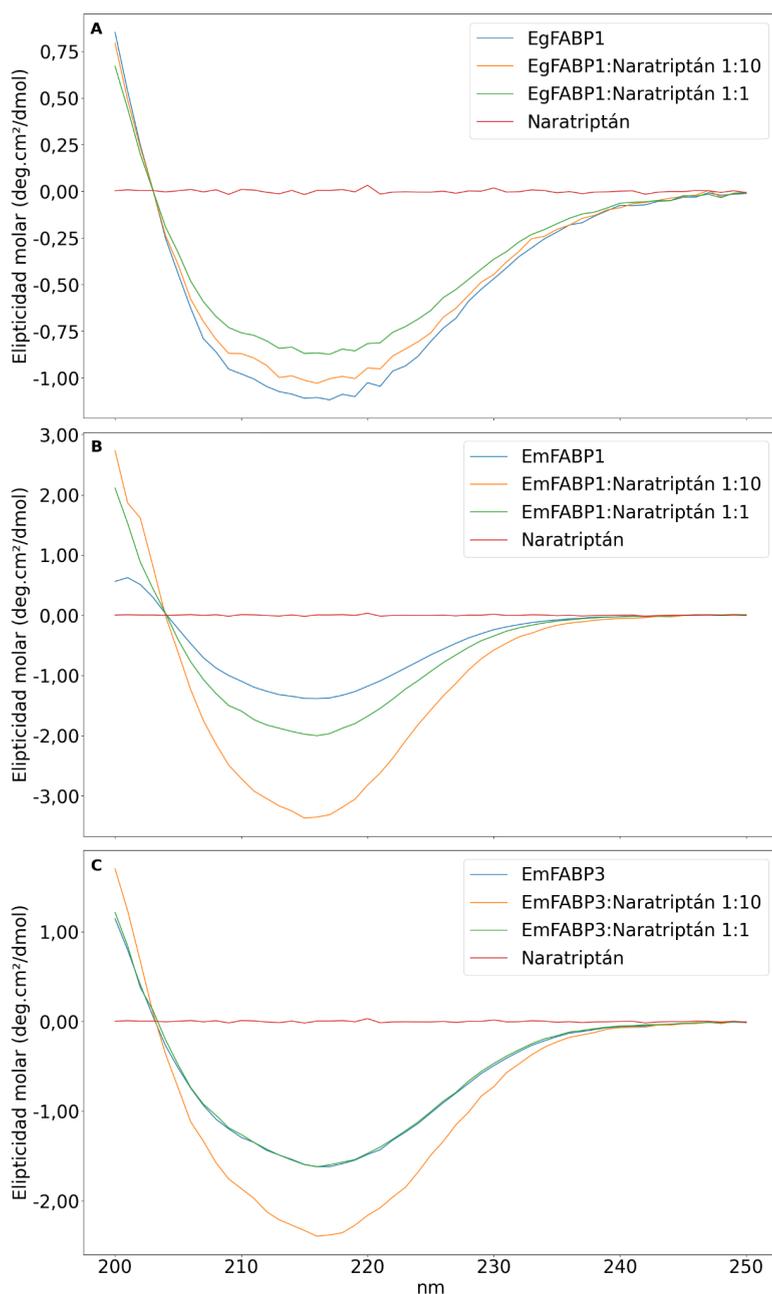


Figura 4.14d. Espectros de *far-UV CD* para Naratriptán en presencia de: (A) **EgFABP1** (8 μM), (B) **EmFABP1** (23 μM), (C) **EmFABP3** (18 μM). La concentración de Naratriptán en ausencia de proteína fue de 50 μM .

REFERENCIAS

Arcon, J. P., Modenutti, C. P., Avendaño, D., Lopez, E. D., Defelipe, L. A., Ambrosio, F. A., Turjanski, A. G., Forli, S., & Marti, M. A. (2019). AutoDock Bias: Improving binding mode prediction and virtual screening using known protein–ligand interactions. *Bioinformatics*, 35(19), 3836–3838. <https://doi.org/10.1093/bioinformatics/btz152>

Bélgamo, J. A., Alberca, L. N., Pórfido, J. L., Romero, F. N. C., Rodriguez, S., Talevi, A., Córscico, B., & Franchini, G. R. (2020). Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets. *Journal of Computer-Aided Molecular Design*, 34(12), 1275–1288. <https://doi.org/10.1007/s10822-020-00352-8>

Cai, H., Wang, T., Yang, Z., Xu, Z., Wang, G., Wang, H.-Y., Zhu, W., & Chen, K. (2017). Combined Virtual Screening and Substructure Search for Discovery of Novel FABP4 Inhibitors. *Journal of Chemical Information and Modeling*, 57(9), 2329–2335. <https://doi.org/10.1021/acs.jcim.7b00364>

Edinburgh Instruments Ltd. (2023). *Raman Scattering in Fluorescence Emission Spectra | Common Errors*. <https://www.edinst.com/us/blog/raman-scattering-blog/>

Floresta, G., Cilibrizzi, A., Abbate, V., Spampinato, A., Zagni, C., & Rescifina, A. (2019a). 3D-QSAR assisted identification of FABP4 inhibitors: An effective scaffold hopping analysis/QSAR evaluation. *Bioorganic Chemistry*, 84, 276–284. <https://doi.org/10.1016/j.bioorg.2018.11.045>

Floresta, G., Cilibrizzi, A., Abbate, V., Spampinato, A., Zagni, C., & Rescifina, A. (2019b). FABP4 inhibitors 3D-QSAR model and isosteric replacement of BMS309403 datasets. *Data in Brief*, 22, 471–483. <https://doi.org/10.1016/j.dib.2018.12.047>

Forli, S., & Olson, A. J. (2012). A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *Journal of Medicinal Chemistry*, 55(2), 623–638. <https://doi.org/10.1021/jm2005145>

He, Y., Dou, H., Gao, D., Wang, T., Zhang, M., Wang, H., & Li, Y. (2019). Identification of new dual FABP4/5 inhibitors based on a naphthalene-1-sulfonamide FABP4 inhibitor. *Bioorganic & Medicinal Chemistry*, 27(19), 115015. <https://doi.org/10.1016/j.bmc.2019.07.031>

Hertzog, A. V., Hellberg, K., Reynolds, J. M., Kruse, A. C., Juhlmann, B. E., Smith, A. J., Sanders, M. A., Ohlendorf, D. H., Suttles, J., & Bernlohr, D. A. (2009).

Identification and Characterization of a Small Molecule Inhibitor of Fatty Acid Binding Proteins. *Journal of Medicinal Chemistry*, 52(19), 6024–6031. <https://doi.org/10.1021/jm900720m>

Jakobsson, E., Alvite, G., Bergfors, T., Esteves, A., & Kleywegt, G. J. (2003). The crystal structure of Echinococcus granulosus fatty-acid-binding protein 1. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1649(1), 40–50. [https://doi.org/10.1016/S1570-9639\(03\)00151-1](https://doi.org/10.1016/S1570-9639(03)00151-1)

Kühne, H., Obst-Sander, U., Kuhn, B., Conte, A., Ceccarelli, S. M., Neidhart, W., Rudolph, M. G., Ottaviani, G., Gasser, R., So, S.-S., Li, S., Zhang, X., Gao, L., & Myers, M. (2016). Design and synthesis of selective, dual fatty acid binding protein 4 and 5 inhibitors. *Bioorganic & Medicinal Chemistry Letters*, 26(20), 5092–5097. <https://doi.org/10.1016/j.bmcl.2016.08.071>

Lakowicz, J. R. (2006). *Principles of fluorescence spectroscopy* (3rd ed). Springer.

López, E. D., Arcon, J. P., Gauto, D. F., Petruk, A. A., Modenutti, C. P., Dumas, V. G., Marti, M. A., & Turjanski, A. G. (2015). WATCLUST: A tool for improving the design of drugs based on protein-water interactions. *Bioinformatics*, 31(22), 3697–3699. <https://doi.org/10.1093/bioinformatics/btv411>

Mariani, V., Biasini, M., Barbato, A., & Schwede, T. (2013). IDDT: a local superposition-free score for comparing protein structures and models using distance difference tests. *Bioinformatics (Oxford, England)*, 29(21), 2722–2728. <https://doi.org/10.1093/bioinformatics/btt473>

Okada, T., Hiromura, M., Otsuka, M., Enomoto, S., & Miyachi, H. (2012). Synthesis of BMS-309403-Related Compounds, Including [¹⁴C]BMS-309403, a Radioligand for Adipocyte Fatty Acid Binding Protein. *Chemical and Pharmaceutical Bulletin*, 60(1), 164–168. <https://doi.org/10.1248/cpb.60.164>

Pharmacognosy group, Uppsala University. (2023). *2D autocorrelations*. Retrieved September 14, 2023, from <https://chemgps.bmc.uu.se/help/dragonx/Autocorrelations.html>

Prada Gori, D. N., Llanos, M. A., Bellera, C. L., Talevi, A., & Alberca, L. N. (2022). iRaPCA and SOMoC: Development and Validation of Web Applications for New Approaches for the Clustering of Small Molecules. *Journal of Chemical Information and Modeling*, 62(12), 2987–2998. <https://doi.org/10.1021/acs.jcim.2c00265>

Quiroga, R., & Villarreal, M. A. (2016). Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLOS ONE*, 11(5), e0155183. <https://doi.org/10.1371/journal.pone.0155183>

Rational Discovery LLC. (2006). *RDKit: Open-Source Cheminformatics and Machine Learning Software, Open-Source Cheminformatics and Machine Learning*. <http://www.rdkit.org/>

Santos-Martins, D., Solis-Vasquez, L., Tillack, A. F., Sanner, M. F., Koch, A., & Forli, S. (2021). Accelerating AUTODOCK 4 with GPUs and Gradient-Based Local Search. *Journal of Chemical Theory and Computation*, 17(2), 1060–1073. <https://doi.org/10.1021/acs.jctc.0c01006>

The AlphaFold Team & Google Colab. (2022). *ColabFold*. <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb#scrollTo=G4yBrceuFbf3>

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. <https://doi.org/10.1002/jcc.21334>

Truchon, J.-F., & Bayly, C. I. (2007). Evaluating Virtual Screening Methods: Good and Bad Metrics for the “Early Recognition” Problem. *Journal of Chemical Information and Modeling*, 47(2), 488–508. <https://doi.org/10.1021/ci600426e>

Volkamer, A., Griewel, A., Grombacher, T., & Rarey, M. (2010). Analyzing the Topology of Active Sites: On the Prediction of Pockets and Subpockets. *Journal of Chemical Information and Modeling*, 50(11), 2041–2052. <https://doi.org/10.1021/ci100241y>

Volkamer, A., Kuhn, D., Grombacher, T., Rippmann, F., & Rarey, M. (2012). Combining Global and Local Measures for Structure-Based Druggability Predictions. *Journal of Chemical Information and Modeling*, 52(2), 360–372. <https://doi.org/10.1021/ci200454v>

Waskom, M. (2021). seaborn: Statistical data visualization. *Journal of Open Source Software*, 6(60), 3021. <https://doi.org/10.21105/joss.03021>

Xu, Q., Huang, L., Liu, J., Ma, L., Chen, T., Chen, J., Peng, F., Cao, D., Yang, Z., Qiu, N., Qiu, J., Wang, G., Liang, X., Peng, A., Xiang, M., Wei, Y., & Chen, L. (2012). Design, synthesis and biological evaluation of thiazole- and indole-based derivatives for the treatment of type II diabetes. *European Journal of Medicinal Chemistry*, 52, 70–81. <https://doi.org/10.1016/j.ejmech.2012.03.006>

CAPÍTULO 5: CONCLUSIONES

5.1 CONCLUSIONES GENERALES

En este trabajo de tesis doctoral se ha llevado adelante un esquema de **CVP** (Kumar & Zhang, 2015), en donde las diferentes bases de datos fueron cribadas por métodos indirectos y directos. Las dianas farmacológicas tenidas en cuenta para el **CVP** fueron las **FABPs** de *E. granulosus*, *E. multilocularis* y *T. solium*, parásitos causantes de la equinococosis quística, equinococosis alveolar y cisticercosis, respectivamente, las cuales son consideradas por la **OMS** como **ETDs**.

Con respecto al **CV** por métodos indirectos, la estrategia de cribado utilizada en esta tesis utilizó el enfoque propuesto en el trabajo de Bélgamo et al. (2020). En dicho trabajo, los autores compilaron compuestos con actividad experimental reportada contra la **AFABP**, basándose en el paradigma de reposicionamiento del blanco (Klug et al., 2016). Bélgamo et al. (2020) utilizaron un conjunto de datos de 258 compuestos, mientras que en esta tesis ese número se extendió a 288, entre **ACTIVOS** e **INACTIVOS** reales. A su vez, la cantidad de *decoys* generada para el cribado retrospectivo en el trabajo recién citado fue de alrededor de 1500, mientras que en este trabajo se utilizaron las herramientas más recientes **LUDe** y **DUDE-Z** para expandir a 7613 y 4965, respectivamente, los compuestos señuelo. Esto permitió la generación y validación de 4000 modelos clasificadores lineales, como así también del ensamble **MIN32**. Se observó que la *performance* del ensamble **MIN32** es menor para la base **LUDe** que para la base **DUDE-Z**, pudiendo esto último deberse a las diferencias en el método de generación de *decoys* de cada herramienta, ya que los compuestos **ACTIVOS** e **INACTIVOS** verdaderos son los mismos en ambas quimiotecas simuladas. Esta observación de que **LUDe** pareciera generar compuestos señuelo más desafiantes para los modelos que **DUDE-Z** ya ha sido reportado en Prada Gori et al. (2022), en donde extienden este análisis a un centenar de conjuntos de datos diferentes.

Se aplicó el ensamble de modelos **MIN32** a las 8 bases de datos ya descritas en esta tesis, contabilizando 435.000 compuestos en total, aproximadamente. De este universo de compuestos, solamente 92 resultaron activos según el ensamble, de los cuales 15 son compuestos en alguna fase de estudio clínico para otra patología. En función de los costos y disponibilidad para su adquisición, los compuestos que finalmente se seleccionaron para su evaluación experimental fueron la Hidroclorotiazida, el Fenticonazol y el Naratriptán.

En cuanto al cribado por métodos directos, se llevó adelante el modelado estructural de las **FABPs** de *E. granulosus*, *E. multilocularis* y *T. solium*. El proceso de modelado, que se llevó adelante mediante ColabFold y **DM**, permitió obtener, hasta donde se conoce, los primeros modelos estructurales teóricos de las casi todas las isoformas de **FABPs** para *Echinococcus spp* y *T. solium* (a excepción del cristal de

EgFABP1). A través de dichos modelos pudo llevarse adelante una campaña de **CV** desde la estructura, tamizando aproximadamente 435.000 compuestos.

El algoritmo de *docking* utilizado en esta tesis fue **AD Bias**, configurando un único sesgo del tipo donador de puentes de hidrógeno en el sitio de unión de las **FABPs**. Este algoritmo fue el que demostró el mejor desempeño tanto en la validación de la función de puntuación como en la validación de la pose experimental. Los resultados parecieran estar en consonancia con la hipótesis de que algunas moléculas de agua en el sitio de unión estabilizan la tríada Arg107, Arg127 y Tyr129, favoreciendo la unión de ciertos tipos de ligandos (Jakobsson et al., 2003). Además, es para destacar la versatilidad que presenta el programa **AD Bias** en cuanto a la cantidad y tipos de sesgos que se pueden implementar, permitiendo al usuario llevar adelante una suerte de “*docking* a medida” para su sistema de estudio. A su vez, la implementación de **AD Bias** en **GPU** permitió acortar los tiempos de **CV** considerablemente, pudiendo cribar grandes quimiotecas de manera eficiente.

Los valores de **PPV** teórico reportados en el capítulo 4 sugieren una gran especificidad del modelo de *docking*, hecho que se refleja en que únicamente fueron clasificados como **ACTIVOS** 98 de algo más de 400.000 compuestos, de los cuales solamente el Montelukast fue seleccionado para la evaluación *in vitro*. De la misma manera que en los modelos **QSAR**, y siendo coincidente con los resultados reportados por Prada Gori et al. (2022), la quimioteca de validación **LUDe** obtuvo valores de **AUCROC** y **PPV** más bajos que **DUDE-Z**, para todos los modelos y algoritmos de *docking* utilizados en esta tesis doctoral.

Con respecto a la validación de la pose experimental, **AD Bias** obtuvo el mejor desempeño entre los algoritmos ensayados, tanto en el conjunto de total de estructuras como en el subconjunto de las **FABPs** de adipocito más la estructura de **EgFABP1**. Es muy importante destacar que las **FABPs** de adipocito son las que comparten el motivo P2 con **EgFABP1** (Jakobsson et al., 2003), razón por la cual se analizó a este subconjunto por separado del total de estructuras.

El desempeño de **AD Bias** evaluado tanto en los experimentos de *re-docking* como de *cross-docking* permiten aseverar que el algoritmo es capaz de reproducir la pose experimental de los distintos ligandos en **EgFABP1**, aunque no es posible generalizar esta afirmación al resto de las estructuras de **FABPs** ensayadas.

Este desempeño pobre en el *docking power* para el conjunto total de estructuras puede deberse, en parte, a las **FABPs** provenientes de diferentes tejidos y especies presentes en el conjunto de datos utilizado para esta validación, ya que es sabido que diferentes isoformas de las **FABPs** de mamíferos unen diferentes tipos de lípidos, incluso mediante diferentes mecanismos (Chmurzyńska, 2006; Haunerland & Spener, 2004). Con esta diversidad estructural en el conjunto de datos se intentó estudiar cuán general es la validez predictiva de la pose experimental sobre las **FABPs**, la cual mostró estar limitada a **EgFABP1**. Por otra parte, el entorno

biológico del receptor en los algoritmos de *docking* dista de ser representativo de la realidad. Rudden & Degiacomi (2021) proponen, en el caso de las proteínas de membrana, simulaciones de **DM** del receptor embebido en una bicapa lipídica para luego tomar distintas *snapshots* del receptor y someterlas a *docking* molecular, mejorando así la *performance* de éste último. En el caso de las **FABPs**, su capacidad de unir lípidos diversos genera un entorno hidrofóbico que tiene influencia en la unión de las moléculas al sitio de unión. Una aproximación interesante podría ser utilizar sesgos para interacciones de tipo hidrofóbicas en **AD Bias**. Para ello, sería necesario contar con información estructural acerca de cuál/es zona/as del sitio de unión podrían favorecerse con este tipo de interacciones, lo cual podría estudiarse, por ejemplo con simulaciones de **DM** en solvente mixto (Arcon et al., 2017).

Con respecto a la parte experimental, se obtuvieron las isoformas recombinantes de **EgFABP1**, **EmFABP1** y **EmFABP3**, cuyos clones de expresión ya se encontraban a disposición en el laboratorio, fruto de los trabajos realizados por Pórfido et al. (2012), y Bélgamo et al. (2020). Estas isoformas fueron utilizadas para llevar adelante ensayos de desplazamiento de fluorescencia de la sonda fluorescente **ANS**. De esta primera ronda de experimentos, se pudo evidenciar el desplazamiento de la sonda **ANS** por el compuesto Hidroclorotiazida, dando valores de **Kdapp** de 2,19, 9,90 y 2,75 μM para **EgFABP1**, **EmFABP1** y **EmFABP3**, respectivamente. Esta diferencia en las **Kdapp** para las diferentes isoformas puede deberse, por un lado, a las diferencias estructurales entre ellas (lo cual era esperable) y también en parte, con el grado de deslipidización que cuenta cada una de ellas, lo cual está relacionado con la diferente afinidad que tienen cada una de estas proteínas por los lípidos provenientes de la purificación en *E. coli* (Bélgamo, Tesis doctoral en curso).

Por el contrario, el resto de los compuestos seleccionados para su evaluación experimental presentaron dificultades para ser ensayados por fluorescencia. En el caso del Montelukast, como se vió en el capítulo 4, se observó un aumento en la emisión de luz en la región del espectro cercana al pico Raman, mientras que la región de emisión del **ANS** permaneció sin cambios. Este comportamiento imposibilita la realización del cálculo de la fluorescencia integrada y en consecuencia, la determinación de las **Kdapp**.

En cuanto al Fenticonazol, se observó que interacciona con la sonda **ANS**, aumentando sus niveles de fluorescencia, aún en ausencia de proteína. Con respecto al Naratriptán, se observó que en ausencia de proteína, se produce un aumento de todo el espectro de emisión, con lo que hace imposible la determinación de constantes de disociación como la llevada a cabo para la Hidroclorotiazida. Sin embargo, cuando el Naratriptán es ensayado en presencia de proteína, se pudo observar un corrimiento hacia la derecha del pico de emisión del **ANS**. Si bien no es posible cuantificar, esto podría ser un indicio de algún tipo de interacción. Dada la

imposibilidad de evidenciar interacción en 3 de 4 candidatos, se procedió a ensayarlos por **CD**.

Se llevaron adelante experimentos de **UV-far CD** para **EgFABP1**, **EmFABP1** y **EmFABP3** con los cuatro compuestos candidatos a reposicionamiento de fármacos. Como se mencionó en el capítulo anterior, el Montelukast estabiliza la región del espectro compatible con las hojas β (220 nm) para **EmFABP1** y **EmFABP3**, mientras que para **EgFABP1** el efecto es el contrario. En cuanto a la Hidroclorotiazida, pareciera estabilizar la región de las hojas β en **EmFABP1**, mientras que para **EmFABP3** este efecto sólo se ve cuando la relación molar proteína:compuesto es 1:10. Para **EgFABP1**, la fracción de hoja β disminuye en presencia del compuesto, sin diferencias entre las relaciones molares 1:1 y 1:10. Resultados similares se encontraron en el **UV-far CD** para el Fenticonazol y el Naratriptán.

A la luz de los resultados heterogéneos entre las diferentes isoformas, el **CD** no resultó ser un método que permita evidenciar interacción **FABP**-compuesto. El mejor ejemplo de esto último es el caso de la Hidroclorotiazida, con la cual se logró cuantificar la interacción con las tres isoformas de **FABPs** por fluorescencia, pero no por **CD**. En consecuencia, con estos resultados no es posible afirmar ni negar la potencial unión de Montelukast, Fenticonazol y Naratriptán a las **FABPs** de cestodos. A su vez, las variaciones descritas en los espectros de **UV-far CD** para cada isoforma evidencian diferencias estructurales entre ellas, algo que se podía inferir a partir de las **Kdapp** reportadas para Hidroclorotiazida en los experimentos de fluorescencia.

Desde un punto de vista funcional, las **FABPs** son difíciles de evaluar experimentalmente. Esto se debe a que su actividad ligadora de lípidos no es algo fácilmente evidenciable, como podría ser, una actividad enzimática, en donde una enzima cataliza una reacción química sobre un compuesto y el producto obtenido puede cuantificarse, por ejemplo, por colorimetría.

Una posibilidad de abordar la dificultad encontrada para evidenciar la unión de los candidatos propuestos a las **FABPs** podría ser la utilización del método de **RMN**. En el trabajo de Hodsdon y Cistola (1997) los autores postulan la interacción de ácidos grasos con la isoforma intestinal humana de **FABP** mediante el cambio en los desplazamientos químicos de ciertos aminoácidos, con respecto a la proteína pura en solución. Otra alternativa podría ser la calorimetría de titulación isotérmica, usada en el trabajo de Capaldi et al. (2009), en donde miden la afinidad del Ácido cólico a la **FABP** ileal del pez cebra. A su vez, Hendrick et al. (2016) obtienen mediante ensayos de desplazamiento por *Surface Plasmon Resonance* (**SPR**) compuestos con afinidad por la **FABP** ileal humana, siendo esta tercera metodología una opción más para evidenciar unión entre los compuestos candidatos y las **FABPs**.

Posible aplicación de los candidatos al reposicionamiento de fármacos

La Hidroclorotiazida es un diurético de la familia de las benzotiadiazinas, cuyo mecanismo de acción es inhibir a los cotransportadores de sodio/cloro del túbulo contorneado distal del riñón, inhibiendo la reabsorción de agua (Herman & Bashir, 2023). Este grupo de diuréticos tiene la capacidad de inhibir, con diferentes grados de potencia, diferentes isoformas de la enzima anhidrasa carbónica humana. Se han reportado casos de actividad antiepiléptica y anticancerígena, correlacionados a la inhibición de determinadas isoformas de anhidrasa carbónica, con lo cual el uso de benzotiadiazinas en combinación con otras drogas para el tratamiento de estas patologías es materia de investigación (Carta & Supuran, 2013).

El Montelukast, por su parte, es un compuesto antagonista de los receptores de leucotrienos, siendo prescrito como profilaxis y tratamiento del asma crónica (Wermuth et al., 2023). Este compuesto resultó ser un caso exitoso de reposicionamiento para el estudio de la Enfermedad de Chagas, obtenido a partir de un **CVP** similar al ensayado en este trabajo de tesis, mostrando actividad *in vitro* tanto contra la proteína blanco como así también frente a diferentes estadios parasitarios (Llanos et al., 2023). Como se mencionó anteriormente, el Montelukast cuenta con propiedades espectrales que impidieron evidenciar si puede ser unido por las **FABPs** de cestodos. En este sentido, es menester evaluarlo mediante alguna técnica experimental que no dependa, en principio, de la absorción y/o emisión de luz.

Por otro lado, el Fenticonazol es un antifúngico de aplicación tópica, el cual actúa por diversos mecanismos, que van desde la inhibición de la secreción de proteasas fúngicas, hasta inhibición de citocromo oxidasa y peroxidasa (Veraldi & Milani, 2008). Dado su vía de administración, aún habiendo evidenciado unión a **FABPs**, este compuesto hubiera requerido de modificaciones químicas posteriores para poder administrarlo por vía oral. En consecuencia, se hubieran perdido ventajas del paradigma de reposicionamiento de fármacos (Temesi et al., 2014), como por ejemplo, los ensayos clínicos y de seguridad temprana, aumentando así los costos en el desarrollo de la nueva terapéutica (Swinney, 2013; Paul & Lewis-Hall, 2013).

Por último, el Naratriptán es un fármaco utilizado para tratar el ataque agudo de migraña o hemicránea, siendo un agonista de los receptores de serotonina 5-HT₁ (Krishna et al., 2017). Al día de la escritura de esta tesis doctoral, no pareciera encontrarse reportado en literatura algún ejemplo de reposicionamiento exitoso para este compuesto. Como ya se mencionó, poner a punto un ensayo adecuado para su evaluación *in vitro* es clave para establecer si interacciona con **FABPs**.

5.2 CONCLUSIÓN FINAL

Como conclusión final de este trabajo de tesis doctoral es de destacar la importancia de las estrategias de cribado *in silico* para encontrar nuevas terapéuticas contra la equinocosis y cisticercosis. Dichas estrategias permitieron la identificación de compuestos candidatos a fármaco en un tiempo considerablemente menor a las metodologías no computacionales y a un costo de inversión considerablemente más bajo, permitiendo, de alguna manera, subsanar la baja inversión privada en **ETDs** y la prevalencia de estas en los sectores de bajos recursos de nuestro país y la región.

REFERENCIAS

Arcon, J. P., Defelipe, L. A., Modenutti, C. P., López, E. D., Alvarez-Garcia, D., Barril, X., Turjanski, A. G., & Martí, M. A. (2017). Molecular Dynamics in Mixed Solvents Reveals Protein–Ligand Interactions, Improves Docking, and Allows Accurate Binding Free Energy Predictions. *Journal of Chemical Information and Modeling*, 57(4), 846–863. <https://doi.org/10.1021/acs.jcim.6b00678>

Bélgamo, J. A., Alberca, L. N., Pórfido, J. L., Romero, F. N. C., Rodriguez, S., Talevi, A., Córscico, B., & Franchini, G. R. (2020). Application of target repositioning and in silico screening to exploit fatty acid binding proteins (FABPs) from *Echinococcus multilocularis* as possible drug targets. *Journal of Computer-Aided Molecular Design*, 34(12), 1275–1288. <https://doi.org/10.1007/s10822-020-00352-8>

Bélgamo, Julián. (Manuscrito en redacción). *Proteínas que unen ácidos grasos (FABPs) de parásitos cestodes: Identificación, caracterización y análisis para su utilización en el desarrollo de nuevas terapéuticas*. Universidad Nacional de La Plata.

Capaldi, S., Saccomani, G., Fessas, D., Signorelli, M., Perduca, M., & Monaco, H. L. (2009). The X-Ray Structure of Zebrafish (*Danio rerio*) Ileal Bile Acid-Binding Protein Reveals the Presence of Binding Sites on the Surface of the Protein Molecule. *Journal of Molecular Biology*, 385(1), 99–116. <https://doi.org/10.1016/j.jmb.2008.10.007>

Carta, F., & Supuran, C. T. (2013). Diuretics with carbonic anhydrase inhibitory action: A patent and literature review (2005 – 2013). *Expert Opinion on Therapeutic Patents*, 23(6), 681–691. <https://doi.org/10.1517/13543776.2013.780598>

Chmurzyńska, A. (2006). The multigene family of fatty acid-binding proteins (FABPs): Function, structure and polymorphism. *Journal of Applied Genetics*, 47(1), 39–48. <https://doi.org/10.1007/BF03194597>

Hauerland, N. H., & Spener, F. (2004). Fatty acid-binding proteins – insights from genetic manipulations. *Progress in Lipid Research*, 43(4), 328–349. <https://doi.org/10.1016/j.plipres.2004.05.001>

Hendrick, A. G., Müller, I., Willems, H., Leonard, P. M., Irving, S., Davenport, R., Ito, T., Reeves, J., Wright, S., Allen, V., Wilkinson, S., Heffron, H., Bazin, R., Turney, J., & Mitchell, P. J. (2016). Identification and Investigation of Novel Binding Fragments in the Fatty Acid Binding Protein 6 (FABP6). *Journal of Medicinal Chemistry*, 59(17), 8094–8102. <https://doi.org/10.1021/acs.jmedchem.6b00869>

Herman, L. L., & Bashir, K. (2023). Hydrochlorothiazide. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK430766/>

Hodsdon, M. E., & Cistola, D. P. (1997). Ligand Binding Alters the Backbone Mobility of Intestinal Fatty Acid-Binding Protein as Monitored by ^{15}N NMR Relaxation and ^1H Exchange. *Biochemistry*, 36(8), 2278–2290. <https://doi.org/10.1021/bi962018l>

Jakobsson, E., Alvite, G., Bergfors, T., Esteves, A., & Kleywegt, G. J. (2003). The crystal structure of Echinococcus granulosus fatty-acid-binding protein 1. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*, 1649(1), 40–50. [https://doi.org/10.1016/S1570-9639\(03\)00151-1](https://doi.org/10.1016/S1570-9639(03)00151-1)

Krishna, K. D., Pradip, S. T., & Navin, S. R. (2017). Preparation and Optimization of Fast Dissolving Film of Naratriptan Hydrochloride. *Recent Patents on Drug Delivery & Formulation*, 11(2). <https://doi.org/10.2174/1872211311666170213105001>

Kumar, A., & Zhang, K. Y. J. (2015). Hierarchical virtual screening approaches in small molecule drug discovery. *Methods*, 71, 26–37. <https://doi.org/10.1016/j.ymeth.2014.07.007>

Llanos, M. A., Alberca, L. N., Ruiz, M. D., Sbaraglini, M. L., Miranda, C., Pino-Martinez, A., Fraccaroli, L., Carrillo, C., Alba Soto, C. D., Gavernet, L., & Talevi, A. (2023). A combined ligand and target-based virtual screening strategy to repurpose drugs as putrescine uptake inhibitors with trypanocidal activity. *Journal of Computer-Aided Molecular Design*, 37(2), 75–90. <https://doi.org/10.1007/s10822-022-00491-0>

Paul, S. M., & Lewis-Hall, F. (2013). Drugs in Search of Diseases. *Science Translational Medicine*, 5(186). <https://doi.org/10.1126/scitranslmed.3004452>

Pórfido, J. L., Alvite, G., Silva, V., Kennedy, M. W., Esteves, A., & Corsico, B. (2012). Direct Interaction between EgFABP1, a Fatty Acid Binding Protein from Echinococcus granulosus, and Phospholipid Membranes. *PLoS Neglected Tropical Diseases*, 6(11), e1893. <https://doi.org/10.1371/journal.pntd.0001893>

Prada Gori, D. N., Alberca, L. N., Rodriguez, S., Alice, J. I., Llanos, M. A., Bellera, C. L., & Talevi, A. (2022). LIDeB Tools: A Latin American resource of freely available, open-source cheminformatics apps. *Artificial Intelligence in the Life Sciences*, 2, 100049. <https://doi.org/10.1016/j.aillsci.2022.100049>

Rudden, L. S. P., & Degiacomi, M. T. (2021). Transmembrane Protein Docking with JabberDock. *Journal of Chemical Information and Modeling*, 61(3), 1493–1499. <https://doi.org/10.1021/acs.jcim.0c01315>

Swinney, D. C. (2013). Phenotypic vs. Target-Based Drug Discovery for First-in-Class Medicines. *Clinical Pharmacology & Therapeutics*, 93(4), 299–301. <https://doi.org/10.1038/clpt.2012.236>

Temesi, G., Bolgár, B., Arany, Á., Szalai, C., Antal, P., & Mátyus, P. (2014). Early repositioning through compound set enrichment analysis: A knowledge-recycling strategy. *Future Medicinal Chemistry*, 6(5), 563–575. <https://doi.org/10.4155/fmc.14.4>

Veraldi, S., & Milani, R. (2008). Topical Fenticonazole in Dermatology and Gynaecology: Current Role in Therapy. *Drugs*, 68(15), 2183–2194. <https://doi.org/10.2165/00003495-200868150-00007>

Wermuth, H. R., Badri, T., & Takov, V. (2023). Montelukast. In *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK459301/>

MATERIAL SUPLEMENTARIO

MATERIAL SUPLEMENTARIO 3

Tabla S3a. Secuencia de aminoácidos de cada una de las proteínas modeladas en esta tesis doctoral (Pórfido et al., 2020; WormBase ParaSite, *Taenia solium*, 2023).

ISOFORMA	SECUENCIA
EgFABP1	MEAF LGTWKMEKSEGFDKIMERLGVDFVTRKMGNLVKPNLIVTDLGGGKYKMRSESTFKTTECSFKLG EKFKVEVTPDSREVTSLITVENGVMKHEQDDKTKVYIERVVEGNELKATVKVDEVVSVRTYSKVA
EgFABP1.2	MEAF LGTWKMEKSEGLEDVMERLKVDMQLRKMATSLKPDLIISDLGDGKYSMRLDSPSIKSECSFKLG EKFKEMTPDKREVSSLITLENGVMKQEQQVEDITVHIERVVEGNELKATGRVDKAVCVLTFTKVA
EgFABP2	MEPFIGTWKMEKSEGFDKIMERLGVDFYFTRKMGNMMPNLIISDLGDGRYNMRSESKFTSEFSFKLG EQFKVEVTPDSREVMSMLTVEDGVLKQEQQVGKDKTTYIDRVVDGNELRATVKADELVCVRTYSRGM
EgFABP3	MDDFLGTWKMTSSEGFDVRMTRLGVDFLTRKAGNAAPVIKLSEVKEGAEPKTYCLKTINAFRSTEVQF RLQEPFLETPDGRKVKTVISLEKDEADGKWVLRQCQVGDKVTDIDRRLDDIDLMRSTIRVEELVSVRM YRRVKD
EgFABP4	MDEFLGSWKLEFNEGLNEILTRLGAGYITRKTNLVSKPVLTFAREHGPDPVEVGEWYSLKTFSTLKSSTI VFRIGYTFDENTTDGRTVRTVFTVDGQVLKQEQQGKTAYITYTREGDLLKTVIKVEDLTCFRNYRRVVQ PPPANGTTPSVSTAKSSEQPTSPTSIEELDNTNSTK
EgFABP5	MEAF LGTWKVDNSGDFHFKFMHFGVPTAIKMNLLVGSKTIFSQKDADTYCIAVSSIKGKTEAARLDQ EFEEACSGGATMKSKITMEGGVMKHIQVGGGETIITRTINGNIMTVVYTVGDLTCSRTRYKRVQ
EmFABP1	MEAF LGTWKLEKSEGFDKIMERLGVDFVTRKMGNLVKPNFVVDLGGGKYKMKSESTFKTTECSFKLG EKFKEMTPDSREVTSLITVEDGVMKHEQDDKTKVYIERVVEGNELKATVKVDEVVSVRTYSKVS
EmFABP2	MEPFIGTWKMEKSEGFDKIMERLGVDFYFTRKMGNMMPNLIISDLGDGRYNMRSESKFTSEFSFKLG EQFKVEVTPDSREVMSMLTVEDGVLKQEQQVGKDKTTYIDRVVDGNELRATVKADELVCVRTYSRGM
EmFABP3	MDDFLGTWKMTSSEGFDVRMTRLGVDFLTRKAGNAAPVIKLSEVKEGAEPKTYCLKTINAFRSTEVQF RLQEPFLETPDGRKVKTVISLEKDEADGKWVLRQCQVGDKVTDIDRRLDDIDLMRSTIRVEELVSVRM YRRVKD
EmFABP4	MDEFLGSWKLEFNEGLNEILTRLGAGYITRKTNLVSKPVLTFAREHGPDPVEVGEWYSLKTFSTLKSSTI VFRIGYTFDENTTDGRTVRTVFTVDGQVLKQEQQGKTAYITYTREGDLLKTVIKVEDLTCFRNYRRVVQ PLPALETTPSMSTAKSTEQPTSPTSIEELDNTNSTK
EmFABP5	MEAF LGTWKVDNSGDFHFKFMHFGVPTAIKMNLLVGSKTIFSQKDADTYCIAVSSIKGKTEATFRLDQE FEEACSGGAPMKSKITMEGGVMKHIQVGRGETIITRTINGNIMTVVYTVGDLTCSRTRYKRVQ
TsFABP1	MEPFIGTWRMEKSEGFDKIMERLGVDFVTRKMGNLMKPSLIVSDLGDGKYSMRSESKFTTEFTFKLG EKFKETTPDSREVTSLITVENGVMKQEQQVGKDKTTYIDRVIEGNELKTTVKVDELVCVRTYVCAA
TsFABP2	MEPFIGTWKMEKSEGFDKIMERLGVDFYFTRKMGNLMKPNLIISDLGDGRYNMRSESKFTSEISFRLGE QFKEVTPDSREVMSMLTVEDGVLKQEQQVGKDKTTYIDRVVEGNELKATVKADELVCVRTYSRGM
TsFABP3	MDNFLGTWKMTSSEGFDVRMARLGVDFLTRKAGNAAPVIKLSEVEDGGDSKTYCLKTISGFRSSEVR FRLQEPFLETTLDGRKVKTVISLEKDEANGRWILRQCQVGDKVTDIDRCLDDIDLMRSTIRVDELVSVRM YRRQKS
TsFABP4	MDEFLGSWKLESNEGLNELLRLGTGYIGRKTNLNATKPVLTFTRENAPGNSAVGEWYSLKTSALKSST CVFRIGYTFDETTADGRTVRTVFTVDGQTLKQEQQGKTAYITYTREGDMLKTVIKLEDLTCFRNYRRVV QPQPANGTTPSVSNSASPPTPPSTPASSNKANKPK
TsFABP5	MEAF LGPWKVESNDDFHFKFMHFGVPTVIKMLQFGSKITFTQKDADTYCSSISSVKGKTEATFRLGE EFEEESPGGTNMKASSKITIEDGVMKHVQVGDSETIITRTINGDIMTVVYTVGDLTCSRTRYKRVK

Tabla S3b. Parámetros utilizados en las corridas de *docking*.

ALGORITMO	PARÁMETROS
Vina/Vina + Vinardo	Exhaustividad=32
AD GPU/AD Hyd/AD Bias	N° de evaluaciones por corrida=2500000 N° máximo de generaciones por corrida=42000 Tamaño inicial de la población=150 Tasa de mutación=2% Tasa de entrecruzamiento=80% Método de optimización local= ADADELTA Número máximo de iteraciones para la optimización local=300 Tasa de optimización local=100 % Tasa de selección para <i>binary tournament</i> = 60 %

Ecuación S3.1. Función de puntuación utilizada por el programa **AD GPU**. ΔG se corresponde con la afinidad predicha para una pose del ligando, ΔH_{vdW} con la contribución entálpica de las interacciones de van der Waals, ΔH_{hbond} con la entalpía asociada a puentes de hidrógeno y ΔH_{elec} que es la entalpía asociada a las interacciones electrostáticas. Por su parte, ΔG_{desolv} representa a la energía libre de desolvatación y ΔS_{tor} es un término que tiene en cuenta la pérdida de entropía del ligando asociada al proceso de unión con el receptor. $W_{vdW}, W_{hb}, W_{el}, W_{desolv}$ y W son términos adimensionales de ajuste determinados experimentalmente con un conjunto de constantes de unión ligando-receptor conocidas. Las constantes A_{ij}, B_{ij} se corresponden con el potencial de Lennard-Jones entre los átomos i, j . Las constantes C_{ij} y D_{ij} se corresponden con el potencial de puente de hidrógeno donador y aceptor entre los átomos i, j . S y V son los parámetros de solvatación y volumen atómico, respectivamente, mientras que σ es una constante fijada en 3.5 Å. La función $s(r_{ij})$ se utiliza para ampliar el mínimo de atracción en los potenciales 12-6 y 12-10. La función $E(t)$ le da direccionalidad al término asociado a puentes de hidrógeno en función del ángulo t . A su vez, q_i y q_j son las cargas atómicas de los átomos i, j , mientras que $\epsilon(r_{ij})$ es la constante dieléctrica del medio en función de la distancia entre los átomos i, j . Por último, N_{tor} es el número de enlaces rotables del ligando, relacionado con el término de pérdida de entropía del ligando (Santos-Martins et al., 2021).

$$\begin{aligned}\Delta G &= \Delta H_{vdW} + \Delta H_{hbond} + \Delta H_{elec} \\ &+ \Delta G_{desolv} + \Delta S_{tor}\end{aligned}$$

(S3.1)

$$\Delta H_{vdW} = W_{vdW} \sum_{i,j} \left(\frac{A_{ij}}{s(r_{ij})^{12}} - \frac{B_{ij}}{s(r_{ij})^6} \right) \quad (I)$$

$$\Delta H_{hbond} = W_{hb} \sum_{i,j} E(t) \left(\frac{C_{ij}}{s(r_{ij})^{12}} - \frac{D_{ij}}{s(r_{ij})^{10}} \right) \quad (II)$$

$$\Delta H_{elec} = W_{el} \sum_{i,j} \left(\frac{q_i q_j}{\varepsilon(r_{ij}) r_{ij}} \right) \quad (III)$$

$$\Delta G_{desolv} = W_{desolv} \sum_{i,j} (S_i V_j + S_j V_i) e^{(-r_{ij}^2/2\sigma^2)} \quad (IV)$$

$$\Delta S_{tor} = W N_{tor} \quad (V)$$

Ecuación S3.2. Función de puntuación de **AD Bias**. El potencial sesgo (V_{bias}) se compone del potencial original de **AD GPU** (V_{ori}) y de un término que tiene en cuenta un potencial energético definido por el usuario (V_{set}), la distancia entre un punto de la grilla (x, y, z) y las coordenadas del centro del sesgo (x_i, y_i, z_i) y el radio de decaimiento del sesgo (r_i) (Arcon et al., 2019).

$$V_{bias} = V_{ori} + V_{set} * e^{(-[(x-x_i)^2 + (y-y_i)^2 + (z-z_i)^2]/r_i^2)}$$

(S3.2)

Ecuación S3.3. Función de puntuación de **AD Hyd**. El valor de ΔG para la pose se corresponde de la suma entre la función de **AD GPU** (ΔG_{lig}) y el valor de afinidad asociado a las aguas unidas al ligando (ΔG_{wat}). Este último término se compone de ΔH_{vdW} , ΔH_{hbond} y ΔS_{desolv} , los cuales se corresponden con la entalpía asociada a interacciones de van der Waals, a puentes de hidrógeno y a la pérdida de entropía por desolvatación, respectivamente entre la molécula de agua W y el átomo j del receptor. K_{wat} es un parámetro constante que describe la pérdida de entropía asociada al desplazamiento de moléculas en el proceso de unión del ligando al receptor. El resto de los parámetros son equivalentes a los descritos en la **Ecuación S3.1** (Forli & Olson, 2012).

$$\Delta G = \Delta G_{lig} + \Delta G_{wat} \quad (\text{S3.3})$$

$$\Delta G_{wat} = \Delta H_{vdW} + \Delta H_{hbond} + \Delta S_{desolv} \quad (I)$$

$$\Delta H_{vdW} = W_{wat} \sum_{w,j} \left(\frac{A_{wj}}{(r_{wj})^{12}} - \frac{B_{wj}}{(r_{wj})^6} \right) \quad (II)$$

$$\Delta H_{hbond} = W_{wat} \sum_{w,j} E(t) \left(\frac{C_{wj}}{s(r_{wj})^{12}} - \frac{D_{wj}}{s(r_{wj})^{10}} \right) \quad (III)$$

$$\Delta S_{desolv} = K_{wat} \quad (IV)$$

Ecuación S3.4. Función de puntuación utilizada en **Vina** y **Vinardo**. El valor de afinidad para la pose i (c) es una sumatoria de las interacciones entre los átomos i, j del ligando y el receptor, la cual es una función dependiente de la distancia entre ellos $f_{t_i t_j}(r_{ij})$. Esta función se describe como el producto entre las contribuciones de tres términos estéricos ($gauss_1$, $gauss_2$ y $repulsion$), un término hidrofóbico, un término asociado a puentes de hidrógeno y otro a pérdida por entropía del ligando y la distancia relativizada a la superficie (d_{ij}), que es la diferencia de la distancia entre los centros de los átomos y sus respectivos radios de van der Waals (R_{t_i}, R_{t_j}). En el caso del término hidrofóbico, el mismo vale 1 cuando la distancia d_{ij} es menor a

0,5 Å, 0 cuando es mayor a 1,5 Å y linealmente interpolable entre dichas distancias. Por parte del término asociado a puentes de hidrógeno, el término vale 1 cuando d_{ij} es menor a -0,7 Å, 0 cuando es mayor a 0 Å y linealmente interpolable entre dichas distancias. Finalmente, la contribución de los enlaces rotables es un valor constante (0,0585). Para el caso de **Vinardo**, la forma de la función de puntuación es la misma, mientras que los parámetros asociados a los pesos, a los puntos de corte en las distancias y en algunos parámetros atómicos difieren (Trott & Olson, 2010; Quiroga & Villarreal, 2016).

$$c = \sum_{i < j} f_{t_i t_j}(r_{ij}) \quad (\text{S3.4})$$

$$d_{ij} = r_{ij} - R_{t_i} - R_{t_j} \quad (\text{I})$$

$$f_{t_i t_j}(r_{ij}) \equiv h_{t_i t_j}(d_{ij}) \quad (\text{II})$$

$$\text{gauss}_1(d) = e^{-(d/0,5 \text{ \AA})^2} \quad (\text{III})$$

$$\text{gauss}_2(d) = e^{-((d-3 \text{ \AA})/2 \text{ \AA})^2} \quad (\text{IV})$$

$$\text{repulsion}(d) = \begin{cases} d^2, & \text{si } d < 0 \\ 0, & \text{si } d \geq 0 \end{cases} \quad (\text{V})$$

Tabla S3c. Códigos PDB de las estructuras utilizadas en la validación de la pose experimental.

CÓDIGO PDB			
1FDQ	2QO4	3P6F	4TJZ
1HMR	3AKM	3P6G	4TKB
1HMT	3FR2	3P6H	5HZ9
1ICM	3FR4	3PPT	5URA
1O8V	3FR5	3WBG	5XNA
1TOU	3HK1	4AZP	6MP4
1TOW	3JSQ	4BVM	
2NNQ	3P6C	4LKT	
2QM9	3P6D	4NNS	

Tabla S3d. Composición de las soluciones utilizadas en esta tesis doctoral.

REACTIVO	COMPOSICIÓN
Gel SDS PAGE 15%	Solución acrilamida-bisacrilamida 30% Resolving buffer Stacking buffer Agua MilliQ Persulfato de Amonio (APS) al 10% Tetrametiletilendiamina (TEMED)
Solución acrilamida-bisacrilamida 30%	30 gr acrilamida 0,8 gr bis-acrilamida 60 mL de agua MilliQ Una vez homogénea, llevar a 100 mL finales con agua MilliQ
Buffer de resolución	19,5 gr Tris 0,4 gr SDS 60 mL de agua bidestilada Llevar a pH=8,8 y a 100 mL finales con agua bidestilada
Buffer de apilamiento	3,02 gr Tris 0,2 gr SDS 20 mL de agua bidestilada Llevar a pH=6,8 y a 50 mL finales con agua bidestilada

Persulfato de Amonio (APS) al 10%	10 mg APS 100 uL de agua bidestilada
Buffer de carga 5X	380 mg Tris 5 mL glicerol 0,1 gr SDS 2,5 mL β -mercaptoetanol 1 mg Azul de Bromofenol Llevar a 10 mL con agua bidestilada
Buffer de corrida	100 mL Buffer LAEMMLI 10 mL SDS al 10 % Llevar a 1000 mL con agua bidestilada
Buffer LAEMMLI	30,3 gr Tris 142,6 gr Glicina 500 mL de agua bidestilada Esperar a disolución total y llevar a 1000 mL con agua bidestilada
Buffer de unión	Tris 20 mM NaCl 500 mM Imidazol 5 mM NaN ₃ 1 mM pH 7,4
Buffer de lavado	Tris 20 mM NaCl 500 mM Imidazol 25 mM NaN ₃ 1 mM pH 7,4
Buffer de elución	Tris 20 mM NaCl 500 mM Imidazol 250 mM NaN ₃ 1 mM pH 7,4
Buffer PBS	NaCl 137 mM KCl 2,7 mM Na ₂ HPO ₄ 10 mM Na ₂ HPO ₄ 1,8 mM
Medio LB	NaCl 10 gr Tryptona 10 gr Levadura 5 gr Llevar a 1000 mL con agua bidestilada

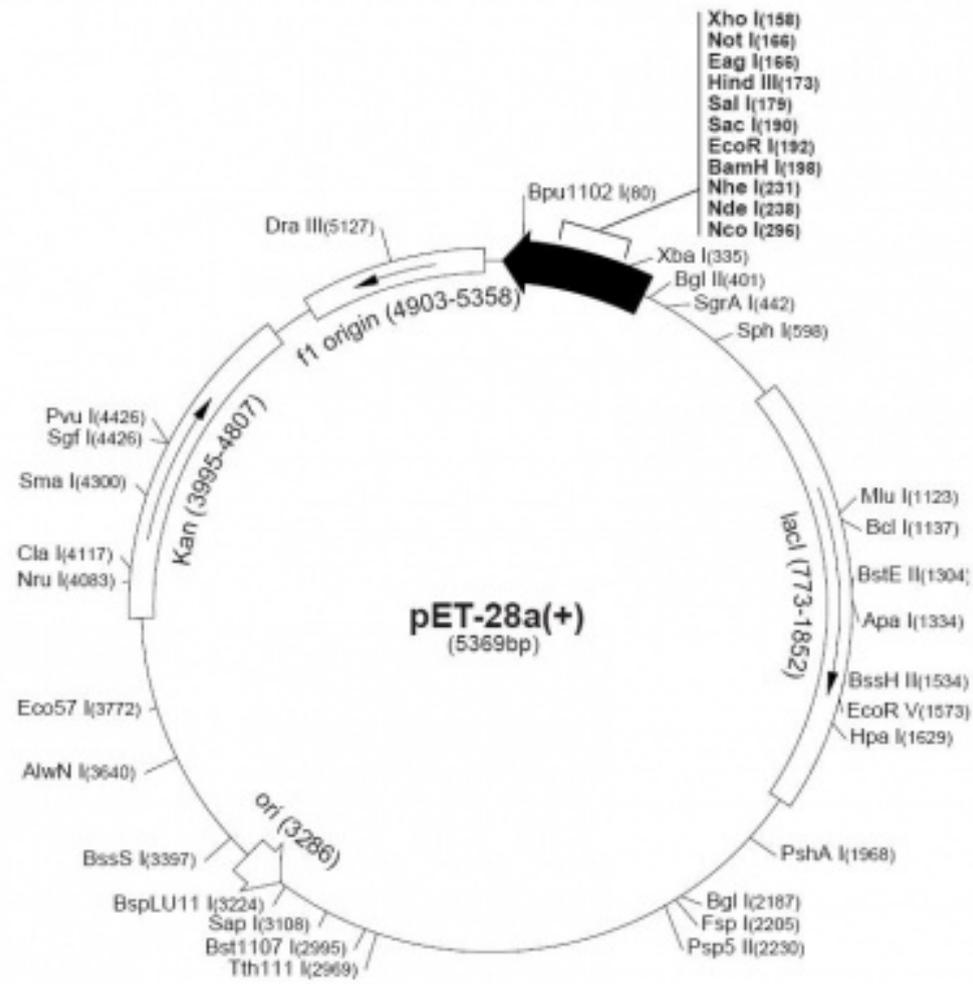


Figura S3a. Vector de expresión pET-28a.



Figura S3b. Vector de expresión pET-11b.

Formalismo S3. Considerando los equilibrios planteados en el capítulo 3, se pueden plantear 5 expresiones (I – V) para este sistema:



+

B

↕

PB

$$Ka = \frac{P^*A}{PA} \text{ (I)} ; Kb = \frac{P^*B}{PB} \text{ (II)}$$

$$P_0 = P + PA + PB \text{ (III)} ; A_0 = A + PA \text{ (IV)}$$

$$B_0 = B + PB \text{ (V)}$$

donde, P_0 , P , PA y PB se corresponden con la concentración de proteína total, libre, y unida con los ligandos A o B , respectivamente. Si se considera al ligando A como la sonda a desplazar (en este caso, **ANS**), A_0 , A y PA se corresponden con la concentración total de sonda, de sonda libre y unida a proteína. Finalmente, B_0 , B y PB se corresponden con la concentración total del ligando de interés, del ligando libre en solución y unido a proteína, respectivamente. Por otra parte, Ka y Kb se corresponden con las constantes de disociación de los ligandos A y B .

Partiendo de la expresión (III), se pueden reemplazar P y PB por las expresiones equivalentes (I) y (II), obteniendo:

$$P_0 = \frac{Ka^* PA}{A} + PA + \frac{Ka^* PA}{A} * \frac{B}{Kb} \text{ (VI)}$$

Reordenando por factor común la expresión (VI) y posteriormente despejando PA , se llega a:

$$P_0 = PA * \left[1 + \frac{Ka}{A} * \left(1 + \frac{B}{Kb} \right) \right] \quad (VII)$$

$$PA = \frac{P_0}{\left[1 + \frac{Ka}{A} * \left(1 + \frac{B}{Kb} \right) \right]} \quad (VIII)$$

Por propiedad distributiva en el denominador en (VIII) y posterior división en tanto a denominador como a numerador por el factor $\left(1 + \frac{Ka}{A} \right)$, se obtiene la expresión X:

$$PA = \frac{P_0}{\left[\left(1 + \frac{Ka}{A} \right) + \left(\frac{Ka * B}{A * Kb} \right) \right]} \quad (IX)$$

$$PA = \frac{\frac{P_0}{\left(1 + \frac{Ka}{A} \right)}}{\left[1 + B * \frac{\frac{Ka}{A * Kb}}{\left(1 + \frac{Ka}{A} \right)} \right]} \quad (X)$$

Finalmente, obtenemos la expresión del modelo (XI), en donde se encuentra la constante aparente de disociación del ligando ensayado ($Kapp$) (XII):

$$PA = \frac{\frac{P_0}{\left(1 + \frac{Ka}{A} \right)}}{\left[1 + \frac{B}{Kapp} \right]} \quad (XI)$$

$$Kapp = \frac{\left(1 + \frac{Ka}{A} \right)}{\frac{Ka}{A * Kb}} \quad (XII)$$

De (XII), es sencillo reordenar términos y llegar a una expresión más compacta (XIII):

$$Kapp = Kb * \left(1 + \frac{A}{Ka} \right) \quad (XIII)$$

Como se puede observar en (XIII), K_{app} tiende al valor de K_b cuando la concentración de la sonda a desplazar A tiende a cero, situación que se da cuando los sitios están libres.

De la expresión XI, se puede notar que la cantidad de PA depende de un cociente. En el numerador se tiene una constante que depende de la cantidad de proteína total P_0 y de la cantidad de sonda a desplazar, cantidades constantes definidas al principio del ensayo, que se corresponderá con el valor de máxima fluorescencia. Por otro lado, en el denominador la única variable es la concentración de ligando desplazante B , por lo cual la cantidad de PA depende de dicha cantidad. Consecuentemente, el valor medido de fluorescencia a la concentración B de ligando se corresponderá con la cantidad de complejo PA formado, mientras que el numerador se corresponderá con la fluorescencia máxima (PA formado cuando $B = 0$). Con ello, se adapta la expresión XI en términos de fluorescencia y se llega a la **Ecuación 3.10**:

$$\Delta F = \frac{\Delta F_0}{1 + \frac{B}{K_{dapp}}} + \Delta F_{res} \quad (3.10)$$

MATERIAL SUPLEMENTARIO 4

Tabla S4a. Compuestos químicos ACTIVOS e INACTIVOS reales compilados en esta tesis doctoral.

COMPUESTO	CLASE	CÓDIGO SMILES	CONJUNTO
A1	ACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OCc3ccccc3)c3ccccc23)ccc1F</chem>	Entrenamiento
A2	ACTIVO	<chem>CCC(CC)c1nc2cc(F)c(Cl)cc2c(-c2ccccc2)c1-c1nnn[nH]1</chem>	Entrenamiento
A3	ACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OCCCC3CCCC3)c3ccccc23)cc1F</chem>	Entrenamiento
A4	ACTIVO	<chem>COc1ccc2ccn(S(=O)(=O)c3ccccc3C(=O)O)c2c1</chem>	Entrenamiento
A5	ACTIVO	<chem>COc1c(-c2ccc(O)cc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccc12</chem>	Entrenamiento
A6	ACTIVO	<chem>Cc1cn(S(=O)(=O)c2ccsc2C(=O)O)c2ccccc12</chem>	Entrenamiento
A7	ACTIVO	<chem>O=C(c1nc2cccc(Cl)c2nc1O)N1CCCC1Cc1cc(Cl)cc(Cl)c1</chem>	Entrenamiento
A8	ACTIVO	<chem>COc1cccc2ccn(S(=O)(=O)c3ccsc3C(=O)O)c12</chem>	Entrenamiento
A9	ACTIVO	<chem>O=C(O)c1cccc(NS(=O)(=O)c2ccc(Br)c3ccccc23)c1</chem>	Entrenamiento
A10	ACTIVO	<chem>Cc1ccc2ccn(S(=O)(=O)c3ccsc3C(=O)O)c2c1</chem>	Entrenamiento
A11	ACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OCC3CCCC3)c3ccccc23)ccc1F</chem>	Entrenamiento
A12	ACTIVO	<chem>CC(C)c1nc2ccc(Cl)cc2c(-c2ccccc2)c1C(=O)O</chem>	Entrenamiento
A13	ACTIVO	<chem>CCOC(=O)C(C)C(=O)CCc1cccc(-c2c(C(=O)O)c(C(C)C)nc3ccc(Cl)cc23)c1</chem>	Entrenamiento
A14	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4ccccc4)c12)CCCC3</chem>	Entrenamiento
A15	ACTIVO	<chem>O=C(Nc1cc(C(F)(F)F)ccc1-c1ccccc1)NC1(C(=O)O)CCCC1</chem>	Entrenamiento
A16	ACTIVO	<chem>COc1c(-c2ccncc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
A17	ACTIVO	<chem>Fc1cc2nc(N3CCCC3)c(-c3nnn[nH]3)c(-c3ccccc3)c2cc1Cl</chem>	Entrenamiento
A18	ACTIVO	<chem>COc1c(-c2ccc(O)c2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccc12</chem>	Entrenamiento
A19	ACTIVO	<chem>CCN(CC)c1nc2ccc(C(F)(F)F)cc2c(-c2ccccc2)c1-c1nnn[nH]1</chem>	Entrenamiento
A20	ACTIVO	<chem>COc1ccc(CSc2nc(O)cc(C(F)(F)F)n2)cc1</chem>	Entrenamiento
A21	ACTIVO	<chem>Cc1nc2ccc(Cl)cc2c(-c2ccccc2)c1C(=O)O</chem>	Entrenamiento
A22	ACTIVO	<chem>O=C(O)c1sccc1S(=O)(=O)n1ccc2cccc(F)c21</chem>	Entrenamiento
A23	ACTIVO	<chem>CCCCCCCCOc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
A24	ACTIVO	<chem>O=C(O)c1c(N2CCCC2)nc2c(Cl)cc(Cl)cc2c1-c1ccccc1</chem>	Entrenamiento
A25	ACTIVO	<chem>Cc1cc(Cl)cc2c(-c3ccccc3)c(C(=O)O)c(N3CCCC3)nc12</chem>	Entrenamiento
A26	ACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OCCC3CCCC3)c3ccccc23)cc1F</chem>	Entrenamiento
A27	ACTIVO	<chem>CCOc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
A28	ACTIVO	<chem>COc1c(-c2ccc(F)cc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccc12</chem>	Entrenamiento

A29	ACTIVO	<chem>O=C(O)c1c(N2CCCC2)nc2ccc(Cl)cc2c1-c1cccc1</chem>	Entrenamiento
A30	ACTIVO	<chem>O=C(O)C=Cc1cn(Cc2cccc2F)c2cccc12</chem>	Entrenamiento
A31	ACTIVO	<chem>Cc1cc(C)c(C)c(S(=O)(=O)Nc2ccc(O)c(Sc3nc[nH]n3)c2)c1C</chem>	Entrenamiento
A32	ACTIVO	<chem>CCN(C)c1nc2ccc(Cl)cc2c(-c2cccc2)c1-c1nnn[nH]1</chem>	Entrenamiento
A33	ACTIVO	<chem>O=C(O)Cc1cn(-c2cccc2-c2nc(-c3ccc4cccc4c3)c(-c3cccc3)s2)nn1</chem>	Entrenamiento
A34	ACTIVO	<chem>FC(F)(F)c1ccc2nc(N3CCCC3)c(-c3nnn[nH]3)c(-c3cccc3)c2c1</chem>	Entrenamiento
A35	ACTIVO	<chem>CCN(C)c1nc2c(Cl)cc(Cl)cc2c(-c2cccc2)c1C(=O)O</chem>	Entrenamiento
A36	ACTIVO	<chem>O=C(O)c1cc(S(=O)(=O)n2ccc3ccc(Br)cc32)cs1</chem>	Entrenamiento
A37	ACTIVO	<chem>O=C(O)c1ccc2c3c(n(Cc4cccc4)c2c1)CCCC3</chem>	Entrenamiento
A38	ACTIVO	<chem>COc1ccc(Cl)cc1NCc1cc(=O)n2nc(-c3cccc3)nc2[nH]1</chem>	Entrenamiento
A39	ACTIVO	<chem>CN(Cc1ccc(Cl)cc1)c1nc(O)cc(C(F)(F)F)n1</chem>	Entrenamiento
A40	ACTIVO	<chem>COc1ccc(S(=O)(=O)Nc2cccc2C(=O)O)cc1C(C)(C)C</chem>	Entrenamiento
A41	ACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2cccc2)n(-c2ccc(Br)cc2)n1</chem>	Entrenamiento
A42	ACTIVO	<chem>O=C(O)COc1ccc(-c2cccc2-c2cc(-c3cccc3)c(-c3cccc3)o2)c1</chem>	Entrenamiento
A43	ACTIVO	<chem>CCc1c(-c2cccc2)c(-c2cccc2)nn1-c1cccc1-c1ccc(OCC(=O)O)c1</chem>	Entrenamiento
A44	ACTIVO	<chem>CCc1c(-c2cccc2)c(-c2cccc2)nn1-c1cccc1-c1ccc(OCC(=O)O)cc1</chem>	Entrenamiento
A45	ACTIVO	<chem>O=C(Nc1cc(Cl)c(F)cc1-c1cccc1)NC1(C(=O)O)CCCC1</chem>	Entrenamiento
A46	ACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2ccc(Cl)cc2)n(-c2cccc2)n1</chem>	Entrenamiento
A47	ACTIVO	<chem>O=C(O)CCc1ccc(O)c(Cc2cccc2)c1</chem>	Entrenamiento
A48	ACTIVO	<chem>O=C(O)Cc1cn(-c2cccc2-c2nc(-c3ccc(Cl)cc3)c(-c3cccc3)s2)nn1</chem>	Entrenamiento
A49	ACTIVO	<chem>O=C(Nc1cc(Cl)ccc1-c1cccc1)NC1(C(=O)O)CCCC1</chem>	Entrenamiento
A50	ACTIVO	<chem>Cn1c(-c2cccc2-c2cccc(NCC(=O)O)c2)nc(-c2cccc2)c1-c1cccc1</chem>	Entrenamiento
A51	ACTIVO	<chem>O=C(O)C1C2CC=CC2c2cc(Cl)cc3c2N1CC1CC=CC31</chem>	Entrenamiento
A52	ACTIVO	<chem>COc1ccc(C(C)C)cc1S(=O)(=O)Nc1cccc(C(=O)O)c1</chem>	Entrenamiento
A53	ACTIVO	<chem>CCc1c(-c2cccc2-c2cccc(OCC(=O)O)c2)oc(-c2cccc2)c1-c1cccc1</chem>	Entrenamiento
A54	ACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2cccc2)n(-c2cccc2Cl)n1</chem>	Entrenamiento
A55	ACTIVO	<chem>CCc1c(-c2cccc2)c(-c2cccc2)nn1-c1cccc1-c1ccc(OC(CC)C(=O)O)c1</chem>	Entrenamiento
A56	ACTIVO	<chem>O=S(=O)(Nc1ccc(O)c(Sc2nc[nH]n2)c1)c1cccc2cccc12</chem>	Entrenamiento
A57	ACTIVO	<chem>CCc1c(-c2cccc2)c(-c2cccc2)nn1-c1cccc1-c1ccc(OC(C)C(=O)O)c1</chem>	Entrenamiento
A58	ACTIVO	<chem>COc1ccc(C(C)C)cc1S(=O)(=O)Nc1cccc(C(=O)O)c1</chem>	Entrenamiento
A59	ACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2ccc(Cl)cc2)n1</chem>	Entrenamiento
A60	ACTIVO	<chem>COc1ccc(C(C)C)cc1S(=O)(=O)Nc1cccc1C(=O)O</chem>	Entrenamiento
A61	ACTIVO	<chem>O=C(O)COc1ccc(-c2cccc2-c2cc(-c3cccc3)c(-c3cccc3)[nH]2)c1</chem>	Entrenamiento

A62	ACTIVO	CCCCCCCC=CCCCCCCC(=O)O	Entrenamiento
A63	ACTIVO	O=C(O)CCOc1cccc1-c1cc(-c2cccc2)n(-c2cccc2)n1	Entrenamiento
A64	ACTIVO	Cc1cc(-c2c3c(nc(C4CCCC4)c2-c2nnn[nH]2)CCC(C)C3)ccn1	Entrenamiento
A65	ACTIVO	O=C(O)CNc1cccc(-c2cccc2-c2nc(-c3cccc3)c(-c3cccc3)n2CCF)c1	Entrenamiento
A66	ACTIVO	O=C(O)C1=CC=CC2C1NC(c1cccc1[N+](=O)[O-])C1CC(Sc3cccc3[N+](=O)[O-])C(Cl)C21	Entrenamiento
A67	ACTIVO	Cc1cn(S(=O)(=O)c2csc(C(=O)O)c2)c2cccc12	Entrenamiento
A68	ACTIVO	O=C(O)COc1cccc(-c2cccc2-c2nc(-c3cccc3)c(-c3cccc3)[nH]2)c1	Entrenamiento
A69	ACTIVO	O=C(Nc1cc(Cl)ccc1-c1ccc(F)cc1)NC1(C(=O)O)CCCC1	Entrenamiento
A70	ACTIVO	CCN(CC)c1nc2cc(F)c(Cl)cc2c(-c2cccc2)c1-c1nnn[nH]1	Entrenamiento
A71	ACTIVO	CC(Oc1cccc(-c2cccc2-c2cc(-c3cccc3)n(-c3cccc3)n2)c1)C(=O)O	Entrenamiento
A72	ACTIVO	CCOc1ccc(C(C)C)cc1S(=O)(=O)Nc1cccc(C(=O)O)c1	Entrenamiento
A73	ACTIVO	O=C(Sc1nc(O)cc(C(F)(F)F)n1)N1CCCC1	Entrenamiento
A74	ACTIVO	CC(C)c1cc(C(C)C)c(C(=O)O)c(C(C)C)c1	Entrenamiento
A75	ACTIVO	COc1ccc(Nc2nc(O)cc(C(F)(F)F)n2)cc1	Entrenamiento
A76	ACTIVO	CCC(C)n1c(-c2cccc2-c2cccc(NCC(=O)O)c2)nc(-c2cccc2)c1-c1cccc1	Entrenamiento
A77	ACTIVO	CCn1c(-c2cccc2-c2cccc(NCC(=O)O)c2)nc(-c2cccc2)c1-c1cccc1	Entrenamiento
A78	ACTIVO	CC(C)c1cc(C(C)C)c(S(=O)(=O)O)c(C(C)C)c1	Entrenamiento
A79	ACTIVO	Cc1ccc(C=c2[nH]c(=O)c(=Cc3cccn3-c3cccc(C(=O)O)c3)s2)cc1C	Entrenamiento
A80	ACTIVO	Cn1c(-c2cccc2-c2cccc(OCC(=O)O)c2)nc(-c2cccc2)c1-c1cccc1	Entrenamiento
A81	ACTIVO	O=C(O)CCCOc1cccc1-c1cc(-c2cccc2)n(-c2ccc(Cl)cc2)n1	Entrenamiento
A82	ACTIVO	CCOc1ccc(S(=O)(=O)Nc2cccc(C(=O)O)c2)c(C)c1C	Entrenamiento
A83	ACTIVO	Cc1ccc(Cn2c3c(c4cccc(C(=O)O)c42)CCCC3)cc1	Entrenamiento
A84	ACTIVO	CCCc1cn(Cc2cccc2)c2c(C(=O)O)cccc12	Entrenamiento
A85	ACTIVO	Cc1c(C)n(Cc2cccc2)c2c(C(=O)O)cccc12	Entrenamiento
A86	ACTIVO	O=C(O)c1cccc2c3c(n(Cc4cccc(F)c4)c12)CCCC3	Entrenamiento
A87	ACTIVO	CCCN1c2c(c3cccc(C(=O)O)c31)CCCC2	Entrenamiento
A88	ACTIVO	O=C(O)c1cccc2c3c(n(Cc4ccc(F)cc4)c12)CCCC3	Entrenamiento
A89	ACTIVO	CCCN1c2c(c3cccc(C(=O)O)c31)CCCC2	Entrenamiento
A90	ACTIVO	O=C(O)c1cccc2c3c(n(Cc4ccc(C(F)(F)F)cc4)c12)CCCC3	Entrenamiento
A91	ACTIVO	COc1cccc(Cn2c3c(c4cccc(C(=O)O)c42)CCCC3)c1	Entrenamiento
A92	ACTIVO	O=C(O)c1cccc2c3c(n(Cc4cccc(F)c4)c12)CCCC3	Entrenamiento
A93	ACTIVO	O=C(O)c1cccc2c3c(n(Cc4ccc(C(F)(F)F)cc4)c12)CCCC3	Entrenamiento
A94	ACTIVO	O=C(O)c1cccc2c3c(n(Cc4cccc4F)c12)CCCC3	Entrenamiento
A95	ACTIVO	COc1cccc1Cn1c2c(c3cccc(C(=O)O)c31)CCCC2	Entrenamiento

A96	ACTIVO	<chem>O=C(O)CCCN1C2C=CC=CC2C2C=CC=CC21</chem>	Entrenamiento
A97	ACTIVO	<chem>CCCOc1cc(S(=O)(=O)Nc2ccc(C(=O)O)cc2)ccc1Cl</chem>	Entrenamiento
A98	ACTIVO	<chem>O=C(O)c1cccc1S(=O)(=O)N1C2C=CC=CC2C2C=CC=CC21</chem>	Entrenamiento
A99	ACTIVO	<chem>Cc1cc(C)c2c3c1C1C=CCC1CN3C(C(=O)O)C1CC=CC21</chem>	Entrenamiento
A100	ACTIVO	<chem>O=C(O)c1sccc1S(=O)(=O)N1C2C=CC=CC2C2C=CC=CC21</chem>	Entrenamiento
A101	ACTIVO	<chem>O=C(O)C1C2CC(Sc3cccc3[N+](=O)[O-])C(Cl)C2c2cccc3c2N1CC1CC=CC31</chem>	Entrenamiento
I1	INACTIVO	<chem>O=C(O)COc1cccc1-c1cc(-c2ccc(Cl)cc2)n(-c2cccc2)n1</chem>	Entrenamiento
I2	INACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2ccc(Cl)cc2)n(-c2cccc2)n1</chem>	Entrenamiento
I3	INACTIVO	<chem>O=C(O)CCCCCOc1cccc1-c1cc(-c2ccc(Cl)cc2)n(-c2cccc2)n1</chem>	Entrenamiento
I4	INACTIVO	<chem>Cc1ccc(-c2cc(-c3cccc3OCCCC(=O)O)nn2-c2cccc2)cc1</chem>	Entrenamiento
I5	INACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2ccc(Br)cc2)n(-c2cccc2)n1</chem>	Entrenamiento
I6	INACTIVO	<chem>COc1cccc1-n1nc(-c2cccc2OCCCC(=O)O)cc1-c1cccc1</chem>	Entrenamiento
I7	INACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2cccc2)n(C2CCCC2)n1</chem>	Entrenamiento
I8	INACTIVO	<chem>O=C(O)CN1C2C=CC=CC2C2C=CC=CC21</chem>	Entrenamiento
I9	INACTIVO	<chem>NC(=O)Cc1cn(S(=O)(=O)c2ccsc2C(=O)O)c2cccc12</chem>	Entrenamiento
I10	INACTIVO	<chem>Cc1ccc2c(ccn2S(=O)(=O)c2cccc2C(=O)O)c1</chem>	Entrenamiento
I11	INACTIVO	<chem>COc1cccc2c1ccn2S(=O)(=O)c1cccc1C(=O)O</chem>	Entrenamiento
I12	INACTIVO	<chem>NC(=O)Cc1cn(S(=O)(=O)c2cccc2C(=O)O)c2cccc12</chem>	Entrenamiento
I13	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3cccc32)c(C(C)C)c1</chem>	Entrenamiento
I14	INACTIVO	<chem>Cc1nc2cccc2n1S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I15	INACTIVO	<chem>Cc1ccc2c(c1)ncn2S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I16	INACTIVO	<chem>Cc1ccc2nc(C)n(S(=O)(=O)c3c(C(C)C)cc(C(C)C)cc3C(C)C)c2c1</chem>	Entrenamiento
I17	INACTIVO	<chem>Cc1ccc2c(c1)nc(C)n2S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I18	INACTIVO	<chem>Cc1nc2ccc(Br)cc2n1S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I19	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3cc(Br)ccc32)c(C(C)C)c1</chem>	Entrenamiento
I20	INACTIVO	<chem>Cc1nc2cc(Br)ccc2n1S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I21	INACTIVO	<chem>Cc1nc2ccc(Cl)cc2n1S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I22	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3ccc(Cl)cc32)c(C(C)C)c1</chem>	Entrenamiento
I23	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3cc(Cl)ccc32)c(C(C)C)c1</chem>	Entrenamiento
I24	INACTIVO	<chem>COc1ccc2nc(C)n(S(=O)(=O)c3c(C(C)C)cc(C(C)C)cc3C(C)C)c2c1</chem>	Entrenamiento
I25	INACTIVO	<chem>COc1ccc2ncn(S(=O)(=O)c3c(C(C)C)cc(C(C)C)cc3C(C)C)c2c1</chem>	Entrenamiento
I26	INACTIVO	<chem>COc1ccc2c(c1)ncn2S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I27	INACTIVO	<chem>COc1ccc2c(c1)nc(C)n2S(=O)(=O)c1c(C(C)C)cc(C(C)C)cc1C(C)C</chem>	Entrenamiento
I28	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3ccc([N+](=O)[O-])cc32)c(C(C)C)c1</chem>	Entrenamiento
I29	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3ccc([N+](=O)[O-])ccc32)c(C(C)C)c1</chem>	Entrenamiento

130	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(=O)(=O)n2cnc3ccc(N)cc32)c(C(C)C)c1</chem>	Entrenamiento
131	INACTIVO	<chem>CC(C)c1cc(C(C)C)c(S(N)(=O)=O)c(C(C)C)c1</chem>	Entrenamiento
132	INACTIVO	<chem>O=C(CNc1nc(O)cc(C(F)(F)F)n1)N1CCCC1</chem>	Entrenamiento
133	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2ccccc2)n1</chem>	Entrenamiento
134	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(Nc2ccccc2)n1</chem>	Entrenamiento
135	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2ccccc2)n1</chem>	Entrenamiento
136	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2cccc(Cl)c2)n1</chem>	Entrenamiento
137	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2cccc2Cl)n1</chem>	Entrenamiento
138	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2ccc(-c3ccccc3)cc2)n1</chem>	Entrenamiento
139	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(NCc2ccncc2)n1</chem>	Entrenamiento
140	INACTIVO	<chem>Cc1ccc(CNc2nc(O)cc(C(F)(F)F)n2)cc1</chem>	Entrenamiento
141	INACTIVO	<chem>CN(Cc1ccccc1)c1nc(O)cc(C(F)(F)F)n1</chem>	Entrenamiento
142	INACTIVO	<chem>Oc1cc(C(F)(F)F)nc(CCc2ccccc2)n1</chem>	Entrenamiento
143	INACTIVO	<chem>COc1ccc(CCc2nc(O)cc(C(F)(F)F)n2)cc1</chem>	Entrenamiento
144	INACTIVO	<chem>Cc1cc(O)nc(NCc2ccc(Cl)cc2)n1</chem>	Entrenamiento
145	INACTIVO	<chem>CCc1cc(O)nc(NCc2ccc(Cl)cc2)n1</chem>	Entrenamiento
146	INACTIVO	<chem>Oc1cc(-c2ccccc2)nc(NCc2ccc(Cl)cc2)n1</chem>	Entrenamiento
147	INACTIVO	<chem>Oc1cc(-c2ccccc2F)nc(NCc2ccc(Cl)cc2)n1</chem>	Entrenamiento
148	INACTIVO	<chem>O=C(O)c1ccc2c(c1)c1c(n2Cc2ccccc2)CCCC1</chem>	Entrenamiento
149	INACTIVO	<chem>O=C1CCCc2c1c1cccc(C(=O)O)c1n2Cc1ccccc1</chem>	Entrenamiento
150	INACTIVO	<chem>O=C(c1nc2ccccc2nc1O)N1CCCC1CN1CCCC1</chem>	Entrenamiento
151	INACTIVO	<chem>O=C(c1nc2ccccc2nc1O)N1CCCC1Cn1cccn1</chem>	Entrenamiento
152	INACTIVO	<chem>O=C(OCc1ccccc1)C1CCCN1C(=O)c1nc2ccccc2nc1O</chem>	Entrenamiento
153	INACTIVO	<chem>CC(C)Oc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
154	INACTIVO	<chem>CCC(C)Oc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
155	INACTIVO	<chem>CCCC(C)Oc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
156	INACTIVO	<chem>CCCCCCCCCCCCOc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
157	INACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OC3CCCC3)c3ccccc23)ccc1F</chem>	Entrenamiento
158	INACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OC3CCCCC3)c3ccccc23)ccc1F</chem>	Entrenamiento
159	INACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OCC3CCOCC3)c3ccccc23)ccc1F</chem>	Entrenamiento
160	INACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(Oc3ccccc3)c3ccccc23)ccc1F</chem>	Entrenamiento
161	INACTIVO	<chem>COc1c(-c2cncnc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
162	INACTIVO	<chem>COc1c(-c2ccnnc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento
163	INACTIVO	<chem>COc1c(-c2ccc(CO)cc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccccc12</chem>	Entrenamiento

164	INACTIVO	<chem>COc1ccc(-c2cc(S(=O)(=O)Nc3ccc(F)c(C(=O)O)c3)c3ccccc3c2OC)cc1OC</chem>	Entrenamiento
165	INACTIVO	<chem>COc1ccc(OC)c(S(=O)(=O)Nc2ccccc2C(=O)O)c1</chem>	Entrenamiento
166	INACTIVO	<chem>COc1cc(S(=O)(=O)Nc2ccccc2C(=O)O)c(OC)cc1C</chem>	Entrenamiento
167	INACTIVO	<chem>COc1cc(C)c(C)cc1S(=O)(=O)Nc1ccccc1C(=O)O</chem>	Entrenamiento
168	INACTIVO	<chem>COc1ccc(C)cc1S(=O)(=O)Nc1ccccc1C(=O)O</chem>	Entrenamiento
169	INACTIVO	<chem>COc1ccc(F)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Entrenamiento
170	INACTIVO	<chem>COc1ccc(Cl)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Entrenamiento
171	INACTIVO	<chem>COc1ccc(Br)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Entrenamiento
172	INACTIVO	<chem>COc1cc(C)c(C)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Entrenamiento
173	INACTIVO	<chem>COc1ccc(C)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Entrenamiento
174	INACTIVO	<chem>COc1cc(C)c(Cl)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Entrenamiento
175	INACTIVO	<chem>COc1ccc(OC)c(S(=O)(=O)Nc2ccc(C(=O)O)cc2)c1</chem>	Entrenamiento
176	INACTIVO	<chem>O=C(O)C1C2CC=CC2c2c3c(cc4ccccc24)C2C=CCC2CN31</chem>	Entrenamiento
177	INACTIVO	<chem>CC(CC(=O)O)c1ccc(N)cc1</chem>	Entrenamiento
178	INACTIVO	<chem>CC(CC(=O)O)c1ccc(O)c([N+](=O)[O-])c1</chem>	Entrenamiento
179	INACTIVO	<chem>C=CC(=O)Oc1cc(O)ccc1C(C)CC(=O)O</chem>	Entrenamiento
180	INACTIVO	<chem>O=C(O)CC(c1ccccc1)c1ccc(O)cc1</chem>	Entrenamiento
181	INACTIVO	<chem>Cc1cccc(C(Cc2ccc(O)cc2)C(=O)O)c1</chem>	Entrenamiento
182	INACTIVO	<chem>CCOc1cc(CCC(=O)O)ccc1O</chem>	Entrenamiento
183	INACTIVO	<chem>Cc1cc(O)ccc1C(CC(=O)O)c1ccccc1</chem>	Entrenamiento
184	INACTIVO	<chem>COc1cc(O)ccc1C(CC(=O)O)c1ccc(O)cc1</chem>	Entrenamiento
185	INACTIVO	<chem>O=C(O)C(Cc1ccc(O)cc1)c1ccc(O)cc1</chem>	Entrenamiento
186	INACTIVO	<chem>O=C(O)C(Cc1ccccc1)c1ccc(O)cc1</chem>	Entrenamiento
187	INACTIVO	<chem>O=C(O)CCc1ccc(O)cc1</chem>	Entrenamiento
188	INACTIVO	<chem>COc1ccc(CCC(=O)O)cc1C</chem>	Entrenamiento
189	INACTIVO	<chem>CCc1cc(CCC(=O)O)ccc1OC</chem>	Entrenamiento
190	INACTIVO	<chem>CCc1cc(CCC(=O)O)ccc1O</chem>	Entrenamiento
191	INACTIVO	<chem>Cc1cc(CCC(=O)O)ccc1O</chem>	Entrenamiento
192	INACTIVO	<chem>O=C(O)CCc1ccc2c(c1)OCCO2</chem>	Entrenamiento
193	INACTIVO	<chem>CCc1cc(O)ccc1CCC(=O)O</chem>	Entrenamiento
194	INACTIVO	<chem>CC(C)c1cc(O)ccc1CCC(=O)O</chem>	Entrenamiento
195	INACTIVO	<chem>Cc1cc(O)ccc1CCC(=O)O</chem>	Entrenamiento
196	INACTIVO	<chem>O=C(O)CCc1ccc2c(c1)CC=N2</chem>	Entrenamiento
197	INACTIVO	<chem>c1cc(CSc2nc3ccccc3[nH]2)nc(CSc2nc3ccccc3[nH]2)c1</chem>	Entrenamiento
198	INACTIVO	<chem>COC(=O)c1cc(OC)c(OC)cc1NC(=O)CSCC(=O)O</chem>	Entrenamiento
199	INACTIVO	<chem>Cc1ccc(SC(CC(=O)c2ccc(C)cc2)C(=O)O)cc1</chem>	Entrenamiento

I100	INACTIVO	<chem>O=[N+][[O-]]c1ccc(S(=O)(=O)c2cccc2)cc1OS(=O)(=O)c1cc(Cl)cc(Cl)c1</chem>	Entrenamiento
I101	INACTIVO	<chem>COc1ccc(-c2c(C(F)(F)F)nn(C)c2NC(=O)CSCC(=O)O)cc1</chem>	Entrenamiento
A102	ACTIVO	<chem>S=c1[nH]nc(-c2c(N3CCCC3)nc3ccc(Cl)cc3c2-c2cccc2)o1</chem>	Prueba
A103	ACTIVO	<chem>O=C(O)c1sccc1S(=O)(=O)n1ccc2cc(Br)ccc21</chem>	Prueba
A104	ACTIVO	<chem>COc1c(-c2cccc(N)c2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccc12</chem>	Prueba
A105	ACTIVO	<chem>Cc1cccc2c1ccn2S(=O)(=O)c1ccsc1C(=O)O</chem>	Prueba
A106	ACTIVO	<chem>Cc1c(C(C)C)ccc(S(=O)(=O)Nc2cccc(C(=O)O)c2)c1C</chem>	Prueba
A107	ACTIVO	<chem>O=C(c1nc2cccc2nc1O)N1CCCC1C(O)c1cccc(Cl)c1</chem>	Prueba
A108	ACTIVO	<chem>O=C(O)c1sccc1S(=O)(=O)n1ccc2ccc(F)cc21</chem>	Prueba
A109	ACTIVO	<chem>CCCCCCCCCOc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2ccc12</chem>	Prueba
A110	ACTIVO	<chem>COc1c(-c2ccc(F)c(F)c2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A111	ACTIVO	<chem>Cc1cn(S(=O)(=O)c2cccc2C(=O)O)c2cccc12</chem>	Prueba
A112	ACTIVO	<chem>CCCCCOc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A113	ACTIVO	<chem>O=C(O)c1sccc1S(=O)(=O)n1ccc2cc(F)ccc21</chem>	Prueba
A114	ACTIVO	<chem>Cc1ccc2ccn(S(=O)(=O)c3csc(C(=O)O)c3)c2c1</chem>	Prueba
A115	ACTIVO	<chem>Cc1ccc2c(ccn2S(=O)(=O)c2ccsc2C(=O)O)c1</chem>	Prueba
A116	ACTIVO	<chem>Clc1ccc2nc(N3CCCC3)c(-c3nn[nH]3)c(-c3cccc3)c2c1</chem>	Prueba
A117	ACTIVO	<chem>CCCCCOc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A118	ACTIVO	<chem>COc1cc(C=CC2SC(=S)N(CCC(=O)O)C2=O)ccc1O</chem>	Prueba
A119	ACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OC3CCCCC3)c3cccc23)ccc1F</chem>	Prueba
A120	ACTIVO	<chem>CCOc1ccc(S(=O)(=O)Nc2ccc(C(=O)O)cc2)c2cccc12</chem>	Prueba
A121	ACTIVO	<chem>COc1c(-c2ccc(Cl)c(Cl)c2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A122	ACTIVO	<chem>NC(=O)c1cccc1Cn1c2c(c3cccc(C(=O)O)c31)CCCC2</chem>	Prueba
A123	ACTIVO	<chem>O=C(O)c1cc(NS(=O)(=O)c2ccc(OC3CCCCC3)c3cccc23)ccc1F</chem>	Prueba
A124	ACTIVO	<chem>COc1c(-c2ccoc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A125	ACTIVO	<chem>COc1ccc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A126	ACTIVO	<chem>COc1ccc2ccn(S(=O)(=O)c3ccsc3C(=O)O)c2c1</chem>	Prueba
A127	ACTIVO	<chem>Cc1ccc(-c2cc(-c3cccc3OCCCC(=O)O)nn2-c2cccc2)cc1</chem>	Prueba
A128	ACTIVO	<chem>COc1c(-c2cccc2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A129	ACTIVO	<chem>Cc1c(Cl)cccc1OCc1cc(=O)n2nc(-c3cccc3)nc2[nH]1</chem>	Prueba
A130	ACTIVO	<chem>COc1c(-c2ccc(F)c(Cl)c2)cc(S(=O)(=O)Nc2ccc(F)c(C(=O)O)c2)c2cccc12</chem>	Prueba
A131	ACTIVO	<chem>O=C(O)c1c(C2CC2)nc2ccc(Cl)cc2c1-c1cccc1</chem>	Prueba
A132	ACTIVO	<chem>COc1ccc(S(=O)(=O)Nc2cccc(C(=O)O)c2)c2cccc12</chem>	Prueba

A133	ACTIVO	<chem>c1cc(-c2c3c(nc(C4CCCC4)c2-c2nnn[nH]2)CCCC3)ccn1</chem>	Prueba
A134	ACTIVO	<chem>O=C(CSc1nc(O)cc(C(F)(F)F)n1)N1CCCC1</chem>	Prueba
A135	ACTIVO	<chem>COc1ccc(Sc2nc(O)cc(C(F)(F)F)n2)cc1</chem>	Prueba
A136	ACTIVO	<chem>Cc1cc(-c2c3c(nc(C4CCCC4)c2-c2nnn[nH]2)CCC(C(F)(F)F)C3)ccn1</chem>	Prueba
A137	ACTIVO	<chem>O=C(O)COc1ccc(-c2cccc2-n2cc(-c3cccc3)c(-c3cccc3)n2)cc1</chem>	Prueba
A138	ACTIVO	<chem>CCc1c(-c2cccc2)c(-c2ccc(F)cc2)nn1-c1cccc1-c1ccc(OCC(=O)O)c1</chem>	Prueba
A139	ACTIVO	<chem>O=C(O)CCOc1cccc1-c1cc(-c2ccc(Br)cc2)n(-c2cccc2)n1</chem>	Prueba
A140	ACTIVO	<chem>O=C(O)CNc1ccc(-c2cccc2-c2nc(-c3cccc3)c(-c3cccc3)[nH]2)c1</chem>	Prueba
A141	ACTIVO	<chem>COC(=O)c1sc(-c2cccs2)cc1NC(=O)C=CC(=O)O</chem>	Prueba
A142	ACTIVO	<chem>O=C(O)CCCCOc1cccc1-c1cc(-c2cccc2)n(-c2cccc2)n1</chem>	Prueba
A143	ACTIVO	<chem>CCc1ccc(OC)c(S(=O)(=O)Nc2cccc(C(=O)O)c2)c1</chem>	Prueba
A144	ACTIVO	<chem>Cc1cc(-c2c3c(nc(C4(C)CCCC4)c2-c2nnn[nH]2)CCCC3)ccn1</chem>	Prueba
A145	ACTIVO	<chem>O=C(O)CCCCOc1cccc1-c1nc(-c2cccc2)c(-c2cccc2)o1</chem>	Prueba
A146	ACTIVO	<chem>CCCCCC1CCc2nc(C3(COC)CCCC3)c(-c3nnn[nH]3)c(-c3cccc3)c2C1</chem>	Prueba
A147	ACTIVO	<chem>O=C(O)COc1ccc(-c2nc(-c3cccc3)c(-c3cccc3)o2)c1</chem>	Prueba
A148	ACTIVO	<chem>Cc1cc(C(C)C)cc(S(=O)(=O)Nc2cccc2C(=O)O)c1C</chem>	Prueba
A149	ACTIVO	<chem>CCn1c(-c2cccc2-c2ccc(OCC(=O)O)c2)nc(-c2cccc2)c1-c1ccc(cc1)</chem>	Prueba
A150	ACTIVO	<chem>CC1COc2c3c(cc(F)c2N2CCN(C)CC2)C(=O)C(C(=O)O)=CC31</chem>	Prueba
A151	ACTIVO	<chem>COc1ccc(C(C)(C)C)cc1S(=O)(=O)Nc1ccc(C(=O)O)cc1</chem>	Prueba
A152	ACTIVO	<chem>O=c1cc(COc2cccc(Cl)c2C2CC2)[nH]c2nc(-c3cccc3)nn12</chem>	Prueba
A153	ACTIVO	<chem>Cc1ccc(-c2cc(-c3cccc3OCC(=O)O)nn2-c2cccc2)cc1</chem>	Prueba
A154	ACTIVO	<chem>O=c1[nH]nc(-c2c(N3CCCC3)nc3ccc(Cl)cc3c2-c2cccc2)o1</chem>	Prueba
A155	ACTIVO	<chem>O=C(O)Cc1cn(-c2cccc2-c2nc(-c3cccc3)c(-c3cccc3)s2)nn1</chem>	Prueba
A156	ACTIVO	<chem>COc1cc(C)c(C(C)C)cc1S(=O)(=O)Nc1cccc(C(=O)O)c1</chem>	Prueba
A157	ACTIVO	<chem>Cc1cc(-c2c3c(nc(C4(CO)CCCC4)c2-c2nnn[nH]2)CCCC3)ccn1</chem>	Prueba
A158	ACTIVO	<chem>O=C(Nc1cc(Cl)cc(Cl)c1-c1cccc1)NC1(C(=O)O)CCCC1</chem>	Prueba
A159	ACTIVO	<chem>CC(C)OC(=O)C(C(=O)C(F)(F)F)c1cc(NS(=O)(=O)c2cccc2)ccc1O</chem>	Prueba
A160	ACTIVO	<chem>COc1ccc(S(=O)(=O)Nc2ccc(C(=O)O)cc2)c2cccc12</chem>	Prueba
A161	ACTIVO	<chem>O=C(O)CCCOc1cccc1-c1cc(-c2cccc2)n(-c2cccc(Cl)c2)n1</chem>	Prueba
A162	ACTIVO	<chem>O=C(Nc1cc(Cl)c(Cl)cc1-c1cccc1)NC1(C(=O)O)CCCC1</chem>	Prueba
A163	ACTIVO	<chem>CCCCCC=CCC=CCCCCCCC(=O)O</chem>	Prueba
A164	ACTIVO	<chem>CCc1c(-c2cccc2)c(-c2cccc2)nn1-c1cccc1-c1ccc(C(O)CC(=O)O)c1</chem>	Prueba
A165	ACTIVO	<chem>CCCCCC=CCC=CCC=CCC=CCCC(=O)O</chem>	Prueba
A166	ACTIVO	<chem>COc1ccc(S(=O)(=O)Nc2cccc(C(=O)O)c2)c(C)c1C</chem>	Prueba

A167	ACTIVO	<chem>CCOc1ccc(S(=O)(=O)Nc2cccc(C(=O)O)c2)cc1C(C)(C)C</chem>	Prueba
A168	ACTIVO	<chem>CCCCCCCCCCCCCCCC(=O)O</chem>	Prueba
A169	ACTIVO	<chem>O=C(O)CCc1ccc(O)c(C(=O)c2ccccc2)c1</chem>	Prueba
A170	ACTIVO	<chem>COc1ccc(Cn2c3c(c4cccc(C(=O)O)c42)CCCC3)cc1</chem>	Prueba
A171	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4c(F)cccc4C(F)(F)F)c12)CCCC3</chem>	Prueba
A172	ACTIVO	<chem>NC(=O)c1cccc(Cn2c3c(c4cccc(C(=O)O)c42)CCCC3)c1</chem>	Prueba
A173	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc(C(F)(F)F)c4)c12)CCCC3</chem>	Prueba
A174	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4ccc(F)cc4)c12)CCCC3</chem>	Prueba
A175	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc4F)c12)CCCC3</chem>	Prueba
A176	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc(C(F)(F)F)c4)c12)CCCC3</chem>	Prueba
A177	ACTIVO	<chem>CCc1c(-c2ccc(F)cc2)c(-c2ccccc2)nn1-c1cccc1-c1ccc(OCC(=O)O)c1</chem>	Prueba
A178	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc4C(F)(F)F)c12)CCCC3</chem>	Prueba
A179	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc4)c12)CCCC3=NO</chem>	Prueba
A180	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc4)c12)CCCC3</chem>	Prueba
A181	ACTIVO	<chem>O=C(O)c1cccc2c3c(n(Cc4cccc4C(F)(F)F)c12)CCCC3</chem>	Prueba
A182	ACTIVO	<chem>O=C(O)CSc1cc(NS(=O)(=O)c2ccc(Cl)cc2)c2ccccc2c1O</chem>	Prueba
A183	ACTIVO	<chem>Cc1cc2c3c(c1)C1C=CCC1C(C(=O)O)N3CC1CC=CC21</chem>	Prueba
A184	ACTIVO	<chem>O=C(O)CCN1C2C=CC=CC2C2C=CC=CC21</chem>	Prueba
A185	ACTIVO	<chem>O=C(O)C1C2CC=CC2c2cc(Br)cc3c2N1CC1CC=CC31</chem>	Prueba
A186	ACTIVO	<chem>O=C(O)COc1cccc(-c2ccccc2-n2cc(-c3ccccc3)c(-c3ccccc3)n2)c1</chem>	Prueba
A187	ACTIVO	<chem>O=C(O)CCCN1C2C=CC=CC2C2C=CC=CC21</chem>	Prueba

Lista S4. Ecuaciones matemáticas de cada uno de los modelos involucrados en el ensamble **MIN32**.

MODELO 3963

$$Y = 0,3832 * nAcid - 0,0399 * ATSC8are - 0,0211 * AATS6i - 0,2626 * AATSC8i + 0,0014 * ATSC3dv + 3,5257$$

MODELO 1940

$$Y = 0,3867 * nAcid + 0,2627 * AATSC3dv - 0,0835 * AATSC8s + 0,1366 * n5aHRing - 0,0002 * ATSC2m + 0,2068$$

MODELO 2836

$$Y = 0,2017 * AATS4d + 0,4267 * nAcid + 0,1158 * ATSC3pe - 0,0023 * ATSC8s + 0,2902 * GATS8v - 0,7271$$

MODELO 3392

$$Y = 0,5263 * nAcid + 0,1224 * ATSC3are + 0,2328 * AATS2d - 0,4795 * MATS8are - 0,3280 * GATS8dv - 0,4254$$

MODELO 3555

$$Y = 0,5342 * nAcid + 0,0106 * ATSC3s - 0,3731 * MATS8c - 0,0001 * VR1_A - 0,0084 * AATS8i + 1,5273$$

MODELO 498

$$Y = 0,4899 * nAcid - 0,5219 * ATSC8c + 0,1612 * ATSC3se + 0,2591 * AATS1d - 0,8769$$

MODELO 3520

$$Y = 0,4087 * nAcid - 0,3794 * MATS8se - 0,0313 * SssO + 0,0196 * nBondsA + 0,2184 * nG12FAHRing - 0,0660$$

MODELO 2525

$$Y = 0,4302 * nAcid + 0,2725 * AATS6d - 0,0355 * AATS8s - 0,0023 * ATSC7s + 0,5559 * ATSC7c - 0,4543$$

MODELO 3732

$$Y = 0,4277 * nAcid - 0,0598 * ATSC8se + 0,4692 * MWC02 - 0,2142 * NssO + 0,3363 * GATS8p - 2,3491$$

MODELO 3668

$$Y = 0,4084 * nAcid - 0,3812 * MATS8s + 0,1335 * naHRing + 0,2821 * IC5 - 0,0115 * PEOE_VSA9 - 1,1934$$

MODELO 2808

$$Y = 0,3664 * nAcid - 0,7298 * MATS8pe + 0,4163 * NaaS - 0,4189 * GATS4s - 0,2157 * GATS8are + 0,7976$$

MODELO 2892

$$Y = 0,6062 * AATS4d + 0,0384 * C2SP3 - 0,2143 * NssO - 0,2031 * SdssC - 0,1481 * nFaRing - 1,4184$$

MODELO 527

$$Y = 0,2582 * nAcid + 0,0056 * MPC4 - 0,0118 * AATSC8m + 0,1545 * ATSC1p - 0,0247 * PEOE_VSA2 - 0,0161$$

MODELO 1992

$$Y = 0,2523 * nAcid + 0,0058 * AATS6v + 0,2713 * nG12FRing - 0,0288 * SlogP_VSA11 - 0,1774 * GATS5s - 0,5156$$

MODELO 2346

$$Y = 0,1555 * piPC9 - 0,2763 * n9FHRing - 0,0441 * SssO - 0,3028 * SdssC + 0,0064 * ATSC3s - 0,6410$$

MODELO 3292

$$Y = 0,1756 * piPC8 - 0,3382 * GATS6are + 0,2205 * ATSC3se - 0,2331 * SdssC - 0,0009 * SpAD_Dzp - 0,1587$$

MODELO 1412

$$Y = 0,2095 * nAcid + 0,0004 * TMPC10 - 0,0001 * ATSC8m - 0,0106 * SlogP_VSA2 + 0,2426$$

MODELO 3674

$$Y = -0,1337 * GATS6s - 0,4315 * MATS8se + 0,0231 * ATSC1i + 0,1451 * ATSC3pe - 0,1897 * SdssC + 0,6155$$

MODELO 2689

$$Y = -0,0280 * AATS6i - 0,0498 * AATSC8Z - 0,2416 * GATS5s + 0,1171 * ATSC3are - 0,1611 * SdssC + 5,1128$$

MODELO 3558

$$Y = 0,2178 * nAcid - 0,0682 * AATSC8Z + 0,0782 * C3SP2 + 0,2745 * nG12FARing + 0,2319 * NaaNH + 0,0608$$

MODELO 3896

$$Y = -0,4483 * GATS6are + 0,2837 * AATSC3dv - 0,0377 * AATS4i - 0,2303 * NssO - 0,2625 * n9FRing + 7,1551$$

MODELO 2933

$$Y = 0,0016 * MPC10 - 0,0138 * AATSC8v - 0,0106 * EState_VSA2 - 0,3016 * SdssC - 0,5428 * ATSC0c + 0,6600$$

MODELO 3744

$$Y = 0,4281 * AATS5d + 0,4334 * nG12FHRing - 0,1883 * SdssC - 0,2444 * GATS5are - 0,0159 * SlogP_VSA3 - 0,5487$$

MODELO 3390

$$Y = 0,4446 * GATS8m + 0,0504 * SaasC + 0,3003 * AATSC3dv - 0,3193 * SdssC + 0,1052 * SaaNH - 0,3248$$

MODELO 769

$$Y = 0,0014 * MPC8 - 0,0195 * ATSC8p - 0,0269 * PEOE_VSA2 + 0,0181 * ATSC1i + 0,2083 * AATSC3dv + 0,4909$$

MODELO 2424

$$Y = -0,3069 * GATS6s + 0,3151 * AATSC3dv + 0,1220 * AATSC1v - 0,0013 * ATSC7dv - 0,0869 * SsNH2 + 0,6509$$

MODELO 327

$$Y = -0,0238 * AATS6i - 0,3807 * GATS5se + 0,2217 * AATSC3dv + 0,0017 * ATSC1v - 0,0097 * AATSC8m + 4,5510$$

MODELO 2269

$$Y = 0,0743 * C3SP2 - 0,0005 * ATSC2v + 0,7346 * AATS3d + 0,6063 * CIC5 - 0,0910 * nHBAcc - 1,9953$$

MODELO 3976

$$Y = -0,6048 * ATSC8c - 0,0125 * ATSC7d + 0,0019 * ATSC1v - 0,2355 * GATS5s - 0,0113 * AATSC6v + 0,5113$$

MODELO 3380

$$Y = 0,3658 * AATS2d - 0,5395 * GATS4s - 0,0320 * PEOE_VSA12 - 0,0752 * ATSC8se - 0,0106 * AATSC8v - 0,3636$$

MODELO 1358

$$Y = -0,0001 * ATSC8m - 0,0264 * AATS5i + 0,3326 * AATSC3dv - 0,0721 * ATSC4are - 0,3909 * GATS4s + 5,0729$$

MODELO 1178

$$Y = 0,5866 * AATS4d - 0,0104 * EState_VSA2 + 0,0798 * AATSC1v - 0,0314 * ETA_dBeta - 0,0072 * ATSC8d - 1,2462$$

Tabla S4b. Descriptores moleculares presentes en los modelos del ensamble **MIN32**.

DESCRIPTOR	DESCRIPCION
nAcid	Cantidad de grupos ácidos
VR1_A	Descriptor VR1 de la matriz de adyacencia
AATS1d	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 1 ponderado por electrones sigma
AATS2d	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 2 ponderado por electrones sigma
AATS3d	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 3 ponderado por electrones sigma
AATS4d	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 4 ponderado por electrones sigma
AATS5d	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 5 ponderado por electrones sigma

AATS6d	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 6 ponderado por electrones sigma
AATS8s	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 8 ponderado por estado intrinseco
AATS6v	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 6 ponderado por volumen de VdW
AATS4i	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 4 ponderado por potencial de ionización
AATS5i	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 5 ponderado por potencial de ionización
AATS6i	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 6 ponderado por potencial de ionización
AATS8i	Coeficiente de autocorrelación promediado de Broto-Moreau de orden 8 ponderado por potencial de ionización
ATSC0c	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 0 ponderado por cargas de gasteiger
ATSC7c	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 7 ponderado por cargas de gasteiger
ATSC8c	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por cargas de gasteiger
ATSC3dv	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 3 ponderado por electrones de valencia
ATSC7dv	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 7 ponderado por electrones de valencia
ATSC7d	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 7 ponderado por electrones sigma
ATSC8d	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por electrones sigma
ATSC3s	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 3 ponderado por estado intrinseco
ATSC7s	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 7 ponderado por estado intrinseco
ATSC8s	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por estado intrinseco
ATSC2m	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 2 ponderado por masa
ATSC8m	Coeficiente de autocorrelación centrado de Broto-Moreau de

	orden 8 ponderado por masa
ATSC1v	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 1 ponderado por volumen de VdW
ATSC2v	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 2 ponderado por volumen de VdW
ATSC3se	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 3 ponderado por electronegatividad de sanderson
ATSC8se	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por electronegatividad de sanderson
ATSC3pe	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 3 ponderado por electronegatividad de pauling
ATSC3are	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 3 ponderado por electronegatividad de allred-rocow
ATSC4are	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 4 ponderado por electronegatividad de allred-rocow
ATSC8are	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por electronegatividad de allred-rocow
ATSC1p	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 1 ponderado por polarizabilidad
ATSC8p	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 8 ponderado por polarizabilidad
ATSC1i	Coeficiente de autocorrelación centrado de Broto-Moreau de orden 1 ponderado por potencial de ionización
AATSC3dv	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 3 ponderado por electrones de valencia
AATSC8s	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 8 ponderado por estado intrínseco
AATSC8Z	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 8 ponderado por número atómico
AATSC8m	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 8 ponderado por masa
AATSC1v	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 1 ponderado por volumen de VdW
AATSC6v	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 6 ponderado por volumen de VdW
AATSC8v	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 8 ponderado por volumen de VdW

AATSC8i	Coeficiente de autocorrelación centrado y promediado de Broto-Moreau de orden 8 ponderado por potencial de ionización
MATS8c	Coeficiente de Moran de orden 8 ponderado por cargas de gasteiger
MATS8s	Coeficiente de Moran de orden 8 ponderado por estado intrínseco
MATS8se	Coeficiente de Moran de orden 8 ponderado por electronegatividad de sanderson
MATS8pe	Coeficiente de Moran de orden 8 ponderado por electronegatividad de pauling
MATS8are	Coeficiente de Moran de orden 8 ponderado por electronegatividad de allred-rocow
GATS8dv	Coeficiente de Geary de orden 8 ponderado por electrones de valencia
GATS4s	Coeficiente de Geary de orden 4 ponderado por estado intrínseco
GATS5s	Coeficiente de Geary de orden 5 ponderado por estado intrínseco
GATS6s	Coeficiente de Geary de orden 6 ponderado por estado intrínseco
GATS8m	Coeficiente de Geary de orden 8 ponderado por masa
GATS8v	Coeficiente de Geary de orden 8 ponderado por volumen de VdW
GATS5se	Coeficiente de Geary de orden 5 ponderado por electronegatividad de sanderson
GATS5are	Coeficiente de Geary de orden 5 ponderado por electronegatividad de allred-rocow
GATS6are	Coeficiente de Geary de orden 6 ponderado por electronegatividad de allred-rocow
GATS8are	Coeficiente de Geary de orden 8 ponderado por electronegatividad de allred-rocow
GATS8p	Coeficiente de Geary de orden 8 ponderado por polarizabilidad
SpAD_Dzp	Desviación espectral absoluta de la matriz de Barysz ponderada por polarizabilidad
nBondsA	Número de enlaces aromáticos en la estructura no kekulizada
C3SP2	Número de carbonos sp ² unidos a otros tres carbonos
C2SP3	Número de carbonos sp ³ unidos a otros dos carbonos
NaaNH	Número de grupos NH
NssO	Número de átomos de O
NaaS	Número de átomos de S

SdssC	Suma de carbonos con doble enlace
SaasC	Suma de carbonos unidos a grupos aromáticos aasC
SsNH2	Suma de grupos NH2
SaaNH	Suma de grupos NH
SssO	Suma de átomos de O
ETA_dBeta	ETA delta beta
nHBacc	Número de aceptores de hidrógeno
IC5	Contenido de información de vecinos de orden 5
CIC5	Contenido de información complementaria de orden 5
PEOE_VSA2	Descriptor 2 VSA de carga MOE (-0.30 <= x < -0.25)
PEOE_VSA9	Descriptor 9 VSA de carga MOE (0.05 <= x < 0.10)
PEOE_VSA12	Descriptor 12 VSA de carga MOE (0.20 <= x < 0.25)
SlogP_VSA2	Descriptor 2 VSA de logP MOE (-0.40 <= x < -0.20)
SlogP_VSA3	Descriptor 3 VSA de logP MOE (-0.20 <= x < 0.00)
SlogP_VSA11	Descriptor 11 VSA de logP MOE (0.50 <= x < 0.60)
EState_VSA2	Descriptor 2 VSA de EState VSA (-0.39 <= x < 0.29)
MPC4	Conteo de elementos de orden 4
MPC8	Conteo de elementos de orden 8
MPC10	Conteo de elementos de orden 10
TMPC10	Conteo total de elementos de orden 10
piPC8	Conteo de elementos pi de orden 8 (escala logarítmica)
piPC9	Conteo de elementos pi de orden 9 (escala logarítmica)
naHRing	Conteo de anillos aromáticos heterogéneos
n5aHRing	Conteo de anillos aromáticos heterogéneos de 5 miembros
n9FRing	Conteo de anillos fusionados de nueve miembros
nG12FRing	Conteo de anillos fusionados de 12 miembros o más
n9FHRing	Conteo de anillos heterogéneos fusionados de nueve miembros
nG12FHRing	Conteo de anillos heterogéneos fusionados de 12 miembros o más
nFaRing	Conteo de anillos aromáticos fusionados
nG12FARing	Conteo de anillos alifáticos fusionados de 12 miembros o más
nG12FAHRing	Conteo de anillos heterogéneos alifáticos fusionados de 12 miembros o más

Los descriptores de autocorrelación 2D son descriptores moleculares que describen cómo una propiedad determinada es distribuida a lo largo de la topología de una estructura molecular. El coeficiente de autocorrelación de Broto-Moreau ($ATSkw$) se define en la **Ecuación S4.1**, siendo w la propiedad atómica a evaluar y siendo k el valor de intervalo de comparación entre elementos, es decir, a cuántos enlaces de distancia se encuentran los átomos comparados dentro de la topología. Por otro lado, nSK es el número de átomos pesados y δ_{ij} es el delta de Kronecker ($\delta_{ij}=1$ si $d_{ij}=k$, en caso contrario es cero, siendo d_{ij} la distancia topológica entre dos átomos considerados) (Pharmacognosy group, Uppsala University, 2023).

$$ATSkw = \frac{nSK-1}{nSK} \sum_{i=1}^{nSK} \sum_{j>i} w_i w_j \delta_{ij} \quad (\text{S4.1})$$

Por su parte, el coeficiente de autocorrelación de Moran ($MATSkw$) se define en la **Ecuación S4.2**, siendo w la propiedad atómica a evaluar y siendo k el valor de intervalo de comparación entre elementos. A su vez, $w_{i,j}$ es la propiedad atómica a evaluar en los átomos i, j , \bar{w} es el valor promedio de la propiedad en la molécula, nSK es el número de átomos diferentes al hidrógeno, δ_{ij} es el delta de Kronecker ($\delta_{ij}=1$ si $d_{ij}=k$, en caso contrario es cero, siendo d_{ij} la distancia topológica entre dos átomos considerados) y Δ es la suma de los δ_{ij} (Pharmacognosy group, Uppsala University, 2023).

$$MATSkw = \frac{\frac{1}{\Delta} \sum_{i=1}^{nSK} \sum_{j=1}^{nSK} \delta_{ij} (w_i - \bar{w})(w_j - \bar{w})}{\frac{1}{nSK} \sum_{i=1}^{nSK} (w_i - \bar{w})^2} \quad (\text{S4.2})$$

Finalmente, el coeficiente de autocorrelación de Geary ($GATSkw$), se define en la **Ecuación S4.3**, siendo w la propiedad atómica a evaluar y siendo k el valor de intervalo de comparación entre elementos. Por otro lado, $w_{i,j}$ es la propiedad atómica a evaluar en los átomos i, j , \bar{w} es el valor promedio de la propiedad en la molécula, nSK es el número de átomos diferentes al hidrógeno, δ_{ij} es el delta de Kronecker ($\delta_{ij}=1$ si $d_{ij}=k$, en caso contrario es cero, siendo d_{ij} la distancia

topológica entre dos átomos considerados) y Δ es la suma de los δ_{ij}
 (Pharmacognosy group, Uppsala University, 2023).

$$GATSkw = \frac{\frac{1}{2\Delta} \sum_{i=1}^{nSK} \sum_{j=1}^{nSK} \delta_{ij} (w_i - w_j)^2}{\frac{1}{(nSK-1)} \sum_{i=1}^{nSK} (w_i - \bar{w})^2} \quad (\mathbf{S4.3})$$

REFERENCIAS

Arcon, J. P., Modenutti, C. P., Avendaño, D., Lopez, E. D., Defelipe, L. A., Ambrosio, F. A., Turjanski, A. G., Forli, S., & Marti, M. A. (2019). AutoDock Bias: Improving binding mode prediction and virtual screening using known protein–ligand interactions. *Bioinformatics*, 35(19), 3836–3838. <https://doi.org/10.1093/bioinformatics/btz152>

Forli, S., & Olson, A. J. (2012). A Force Field with Discrete Displaceable Waters and Desolvation Entropy for Hydrated Ligand Docking. *Journal of Medicinal Chemistry*, 55(2), 623–638. <https://doi.org/10.1021/jm2005145>

Pharmacognosy group, Uppsala University. (2023). *2D autocorrelations*. Retrieved September 14, 2023, from <https://chemgps.bmc.uu.se/help/dragonx/Autocorrelations.html>

Pórfido, J. L., Herz, M., Kiss, F., Kamenetzky, L., Brehm, K., Rosenzvit, M. C., Córscico, B., & Franchini, G. R. (2020). Fatty acid-binding proteins in *Echinococcus* spp.: The family has grown. *Parasitology Research*, 119(4), 1401–1408. <https://doi.org/10.1007/s00436-020-06631-5>

Quiroga, R., & Villarreal, M. A. (2016). Vinardo: A Scoring Function Based on Autodock Vina Improves Scoring, Docking, and Virtual Screening. *PLOS ONE*, 11(5), e0155183. <https://doi.org/10.1371/journal.pone.0155183>

Santos-Martins, D., Solis-Vasquez, L., Tillack, A. F., Sanner, M. F., Koch, A., & Forli, S. (2021). Accelerating Auto Dock 4 with GPUs and Gradient-Based Local Search. *Journal of Chemical Theory and Computation*, 17(2), 1060–1073. <https://doi.org/10.1021/acs.jctc.0c01006>

Trott, O., & Olson, A. J. (2009). AutoDock Vina: Improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *Journal of Computational Chemistry*, NA-NA. <https://doi.org/10.1002/jcc.21334>

WormBase ParaSite. (2023). *Taenia solium*. Retrieved September 13, 2023, from https://parasite.wormbase.org/Taenia_solium_prjna170813/Info/Index/